University of Alberta

COST-EFFECTIVE SAMPLING DESIGN FOR LARGE-SCALE AVIAN MONITORING

By

Matthew Carlson

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Conservation Biology

Department of Renewable Resources

Edmonton, Alberta

Spring 2001

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60419-5

Canada

University of Alberta

Library Release Form

Name of Author: Matthew John Carlson

Title of Thesis: Cost-effective Sampling Design for Large-scale Avian Monitoring

Degree: Master of Science

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Matthew C-L

9617 83 Avenue
Edmonton, Alberta T6C 1C1

Date: November 9, 2000

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled COST-EFFECTIVE SAMPLING DESIGN FOR LARGE-SCALE AVIAN MONITORING submitted by MATTHEW CARLSON in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE in CONSERVATION BIOLOGY.

Dr. Fiona Schmiegelow (Supervisor)

Dr. Brad Stelfox

Dr. Vic Adamovicz

Dr. Peter Boxall

November 7, 2000

Abstract

Despite their important roles in biodiversity conservation, large-scale ecological monitoring programs are scarce, in large part due to the difficulty of achieving effective design under fiscal constraints. Using long-term avian monitoring in Alberta's boreal forest as an example, I evaluated several sampling designs for cost-effectiveness. Empirical parameter estimates were used in simulations that estimated power of designs to detect trend in species' populations and community metrics. The ability to detect trend with increased sample effort depended on the monitoring target's variability and how effort was allocated to sampling parameters. Integration of power and cost estimates identified two cost-effective sampling strategies: multiple within-site plots are preferable to multiple within-year surveys for decreasing sample error; and, many sites should be sampled relatively infrequently rather than sampling few sites frequently, although importance of frequent sampling increases for variable targets. To facilitate cost-effective monitoring of biodiversity, this analysis must be extended to additional taxa.

Acknowledgements

My supervisor, Fiona Schmiegelow, deserves many thanks for her support and advice during this thesis work. I am also thankful for the perspectives provided by members of my supervisory committee, Brad Stelfox, Vic Adamowicz, and Peter Boxall. I am grateful to Steve Cumming for his generous assistance with computer programming and statistical modeling issues. Interaction with the Alberta Forest Biodiversity Monitoring Program Technical Committee was also of great help, and I thank its members, especially Dan Farr.

This research greedily used data not collected by the author. Without the hard work of those who collected the data, this thesis would not exist. Many thanks to Breeding Bird Survey volunteers, and to those who have collected data over the years at Calling Lake. Thanks also to those who provided computer code. This research was funded by the Natural Sciences and Engineering Research Council of Canada and the Sustainable Forest Management Network. Additional funding was provided by the Alberta Forest Biodiversity Monitoring Program.

Perhaps even more important than intellectual and financial assistance has been the emotional support I have been fortunate to receive. Family and friends deserve thanks for keeping me sane. Thanks especially to Ann and Kristian for providing a cottage in the mountains. Finally, and most importantly, thank you Kelly. Your wisdom and support made this happen.

Table of Contents

Chapter 1 *General Introduction*

Introduction 1
The Roles of Ecological Monitoring 2
Birds as Monitoring Targets 3
The Importance of Sampling Design 4
Thesis Outline 6

Chapter 2 *Variability in Avian Abundance and Community Metrics*

Introduction 7
The Model 9
Methods 10
Species and Community Metric Selection 10
Estimating Parameters 12
Within-site Temporal Variance and Mean Initial Abundance 12
Between-site Variation 13
Distribution Assumptions 15
Results 16
Species Selection 16
Parameter Estimates 18
Discussion 19
Summary of Results and Implications for Monitoring 19
The Influence of Within-site Effort on Within-site Temporal Variance 19
The Variability of Species 21
The Variability of Community Metrics 22
Methodology 23
Within-site Temporal Variance 23
Between-site Initial Variance 24

Between-site Trend Variance 25
Data Limitations 25
Summary 26

Chapter 3 *The Effect of Sample Effort on Power*

Introduction 28
Methods 30
Monte Carlo Simulations 30
Assumed Data Structure 30
Simulation Process 32
Simulation Scenarios 33
Sensitivity Analysis 35
Modeling the Relationship Between Power and Sample Effort 36
Estimating Rates of Power Gain with Effort 37
Results 38
Power Estimates 38
20 Year Monitoring Programs 38
10 Year Monitoring Programs 38
Sensitivity Analysis 41
Power Models 41
Rates of Power Gain with Effort 42
Discussion 43
Power to Detect Population Trends 43
Power to Detect Community Trends 44
The Effect of Program Length on Power 45
The Effect of Sample Effort on Power 45
Implications of Simulation Assumptions 47
Summary 49

Chapter 4 *Cost-effective Sampling Strategies for Detecting Long-term Avian Trends*

Introduction 50
Methods 51
Estimating Sampling Design Cost 52
Field Costs 52
Aerial Access Cost 53
Estimating Least Expensive Effective Designs 54
Estimating Relative Efficiency of Effort Parameters 55
Results 56
20 Year Monitoring Programs 56
Cost Estimates 56
Least Expensive Effective Designs 56
Relative Efficiency of Effort Parameters 57
10 Year Monitoring Programs 59
Discussion 61
Cost Estimates 61
Cost-effective Monitoring Program Design 62
Least Expensive Effective Designs 62
Relative Efficiency of Effort Parameters 63
Extending the Analysis to Additional Taxa 67
Summary 67

Chapter 5 *General Conclusions*

Introduction 69
General Findings 69
Extending the Analysis to Multiple Taxa 72
Optimal Experimental Design 73
General Implications 75

Literature Cited 77

Appendix 1 Parameter Estimates 84
Appendix 2 Power Model Partial Derivatives 86
Appendix 3 Cost Model Partial Derivatives 87

List of Tables

2.1 Response of species' within-site temporal variance to within-site sample effort 18

2.2 Response of community metrics' within-site temporal variance to within-site sample effort 19

2.3 Mean estimates of variance parameters across selected species and community metrics 19

2.4 Spatial and temporal extend of data sets required to estimate variance parameters 26

3.1 Description and estimation methods for variance parameters 31

3.2 Sampling designs evaluated in power simulations 34

3.3 Results of a sensitivity analysis evaluating the influence of species' parameters on power to detect trend 41

3.4 Coefficients of determination for nonlinear regressions relating sample effort parameters to power 41

3.5 Rates of change in power to detect 20 year trends, with respect to four sample effort parameter 43

3.6 Rates of change in power to detect 10 year trends, with respect to four sample effort parameter 43

4.1 Mean annual cost and percentage of annual cost accountable to aerial access for 20 year programs, for three aerial access cost decline scenarios 56

4.2 Least expensive effective designs for 20 year programs 58

4.3 Relative efficiencies of effort parameters for detecting 20 year trends 59

4.4 Mean annual cost and percentage of annual cost accountable to aerial access for 10 year programs, for three aerial access cost decline scenarios 60

4.5 Least expensive effective designs for 10 year programs 60

4.6 Relative efficiencies of effort parameters for detecting 10 year trends 61

A1.1 Within-site temporal variance estimates 84

A1.2 Between-site initial variance and between-site trend variance estimates 84

A1.3 Mean abundance estimates 85

A2.1 Partial derivatives of the power model for 20 year programs 86

A2.2 Partial derivatives of the power model for 10 year programs 86

A3.1 Partial derivatives of the cost model 87

A3.2 Rates of increase in total program cost with sample effort parameters for 20 year programs 87

A3.3 Rates of increase in total program cost with sample effort parameters for 10 year programs 87

List of Figures

2.1 Generalized design of a large-scale avian monitoring program 8

2.2 Temporal and spatial variability in bird species abundance and community metrics 17

3.1 Response of power to sample effort for 20 year programs 39

3.2 Response of power to number of sites for detecting trends in Black-capped Chickadee abundance over 10 and 20 years 40

3.3 Response of power to sample effort for 10 year programs 40

3.4 Rate of change in power to detect 20 year trends in Black-capped Chickadee abundance with respect to four sample effort parameters, average across designs achieving one of five levels of power 42

4.1 Power to detect 20 year trends in Black-capped Chickadee abundance versus annual program cost 57

4.2 Minimum effective monitoring program cost versus within-site temporal variance for 20 year programs 58

Chapter 1

General Introduction

Introduction

Anthropogenic disturbance of the environment is extensive and increasing due to expanding human populations and increased resource consumption (Tyler Miller 1993). In Alberta, the structure of the forested region is being drastically altered by rapid development (Thomas 1998). Current knowledge is insufficient to predict the effects of this disturbance on biodiversity (Walters and Holling 1990). The combination of large-scale disturbance and incomplete knowledge impedes effective conservation of biodiversity, necessitating monitoring of ecosystems (Furness et al. 1993). Monitoring can facilitate responsible management of ecosystems by improving the information available to managers (Stork et al. 1995). Despite this well recognized need, however, funding to increase monitoring activities is limited (e.g. Burbidge 1991), and ecological monitoring programs remain rare (Schneider et al. 1999). Currently, no adequate program exists to effectively monitor the biodiversity of Alberta's forested regions (Lee and Hanus 1998), motivating a proposal to develop the Alberta Forest Biodiversity Monitoring Program (Farr et al. 1999). Faced with extensive data requirements and limited resources, cost-effective monitoring is essential for the success of this program, and other large-scale ecological monitoring programs (Committee on Environment and Natural Resources 1997).

This thesis investigates cost-effective large-scale ecological monitoring, using avian communities in Alberta's forested region as a case study. In this chapter, context for the research is provided by exploring the roles of large-scale ecological monitoring, the suitability of avian communities to monitoring, and the importance of sampling design to achieving an effective monitoring program. This chapter concludes with an overview of the thesis.

The Roles of Ecological Monitoring

Ecological monitoring is the repeated measurement of an ecosystem component over time and space, with the intent of detecting change (Stork and Samways 1995). Ecological monitoring provides a basis for managing biological resources and effectively conserving ecosystems and populations (Spellerberg 1991). It facilitates adaptive management by providing an information feedback loop to assess the impact of management on ecosystems (Bernstein and Zalinski 1983, Heywood and Baste 1995). Monitoring assists in the setting of conservation priorities by highlighting degraded and threatened ecosystem components (Stork and Samways 1995). Similarly, monitoring can indicate deficiencies in current knowledge, suggesting where more intensive research is needed (Burbidge 1991). Perhaps most importantly, monitoring informs the public on the status of ecosystems (Committee on Environment and Natural Resources 1997). In a democracy, where well-informed citizens are critical to effective governance, this facilitates the formation of appropriate policy.

The vast majority of ecological studies are less than two years in length and confined to small areas (Committee on Environment and Natural Resources 1996). Results from such studies often cannot be used to solve applied ecological problems, where long-term processes operating over large spatial scales are usually of interest (Pimm 1991). Ecological monitoring, if it operates over large spatial and temporal scales and remains consistent in protocol, can assist in bridging the gap between research and applied science. Due to the short-term nature of most research, natural temporal ecosystem patterns are poorly understood. Long-term monitoring contributes to this knowledge, in so doing helping to decipher natural patterns from patterns due to anthropogenic disturbance (Spellerberg 1991, Samways et al. 1995), and providing a background for the interpretation of results from short-term studies (Committee on Environment and Natural Resources 1996). The limited spatial scale of most research inhibits extending results to regional scales due to the large spatial variation exhibited by most ecological quantities (Dixon et al. 1998). By achieving more complete coverage, spatially extensive monitoring programs facilitate the generalization of results to more policy relevant spatial scales. This thesis explores the design of long-term, spatially extensive (hereafter termed

large-scale) monitoring programs. The assumed purpose of monitoring is detection of temporal trend in population size or community structure. The existence of trend, defined here as long-term change in the mean, lies at the heart of many ecological issues (Dixon et al. 1998).

The importance of ecological monitoring is well recognized. The British Ecological Society identified long-term ecological monitoring as one of its main research priorities (Hassell 1989). The need to monitor ecosystems is also identified in policy at international, national, and provincial levels. Article 7 of the Convention on Biodiversity requires signatories to monitor biodiversity (Johnson 1993). As a signatory to the Convention, Canada is committed to monitoring biodiversity. This commitment is expressed in the Canadian Biodiversity Strategy (Minister of Supply and Services 1995) and the National Forest Strategy (National Resources Canada 1998). Provincially, the government of Alberta recognizes the importance of ecosystem monitoring, and the need for its improvement, in the Alberta Forest Legacy document (Alberta Environmental Protection 1998) and the Alberta Forest Conservation Strategy (Alberta Forest Conservation Strategy Steering Committee 1997).

Natural resource managers in Alberta are in the process of implementing the Alberta Forest Biodiversity Monitoring Program. The goal of the program is to detect changes in biodiversity of Alberta's forests, and thereby measure success towards the broader goal of sustainable development (Schneider et al. 1999). The current emphasis is on detection of trends in ecological and administrative regions, such as natural subregions (Alberta Environmental Protection 1994) and Forest Management Agreement areas. This will be achieved using field sampling and remote sensing to sample a selection of terrestrial and aquatic plant and animal species, and their habitats, at sites systematically or randomly located across Alberta's forested region.

Birds as Monitoring Targets

The extent of biodiversity restricts monitoring to a subset of existing biota. Bird communities are one frequently monitored component (e.g. Dunn et al. 1993), and multiple characteristics make them a suitable target. Due to their relatively high place on

the food chain, birds are sensitive to environmental change (Baillie 1991). Avian response to disturbance incorporates numerous disturbance types due to bird species' diverse habitat requirements, and is rapid due to bird species' relatively short generation times (Steele et al. 1984). Intensive research of bird communities has resulted in a reasonable understanding of avian biology, and this knowledge assists the interpretation of monitoring data (Furness et al. 1993, Moors 1993). Birds are relatively easy to identify and bird systematics are well developed, facilitating accurate data collection (Furness et al. 1993). Standard sampling protocols are established for bird communities (e.g. Bibby et al. 1992) and, relative to other taxa, these protocols efficiently provide large amounts of data (e.g. Croonquist and Brooks 1991). Finally, birds are perhaps the most understood and admired biodiversity component by the general public. This makes bird communities an expedient medium through which to communicate the status of the environment (Furness et al. 1993).

The Importance of Sampling Design

In large-scale monitoring, the population of interest is too large to census. Instead, the population is sampled in order to estimate population parameters. The ability of a monitoring program to fulfill desired roles depends on the quality of the parameter estimates it facilitates. Although statistical power will be introduced as a more objective measure of data quality in Chapter 3, for now, quality can be understood as how well an estimate represents the true parameter value. The quality of a parameter estimate is influenced by how a population is sampled (Thompson et al. 1998), implying the utility of a monitoring program is dependent on whether an appropriate sampling design is employed (Dixon et al. 1998).

A sampling design is a prescription for selecting parts of a population for measurement (Scott 1998). The paucity of large-scale ecological monitoring programs may be attributed, in part, to challenges faced in achieving an appropriate sampling design. The Environmental Monitoring and Assessment Program (United States Environmental Protection Agency 2000), perhaps the most ambitious large-scale ecological monitoring program to date, was determined through review to be unlikely to

meet its goal of monitoring status and trends of USA's ecological resources (National Research Council 1995). Among the deficiencies highlighted by the National Research Council was insufficient sampling intensity to achieve program goals.

An effective sampling design must be statistically credible and efficient (Hinds 1984). To be statistically credible, the sampling design must provide data that facilitate reliable estimation of parameters. Of interest here is estimation of trend. Two factors influence the quality of the parameter estimate: bias and precision (Thompson et al. 1998). Bias is the difference between the expected value of an estimator and the true parameter value. The magnitude of this difference, which should be minimized, is influenced by how data are collected, and how parts of the population are selected for measurement. Sampling designs considered in this thesis use counts of birds to derive trend estimates. This requires the assumption that counts represent constant proportions of true bird abundance (Lewis and Gould 2000). This assumption is difficult, if not impossible, to validate, but it is accepted here. Similarly, using a selection of sites to make inferences to a population assumes sampled areas are representative of areas not sampled (Gibbs and Melvin 1997). Acceptance of this assumption is justified by the use of random sampling in the designs considered in this research.

Precision refers to variation among repeated samples. Precision should also be minimized to achieve good parameter estimates. Biological populations vary temporally and spatially due to factors such as weather, habitat composition, and predator-prey interactions. Therefore, some natural variation in a monitored quantity is expected. However, such variation is increased by sampling variation, due to random selection of sampling units and incomplete counts within each unit (Thompson et al. 1998). Variation due to random selection of sampling units decreases as more units are selected, and variation due to incomplete counts decreases as more effort is exerted in collecting the counts. Therefore, increasing sampling effort decreases sampling variation, and increases precision. This, in turn, facilitates more reliable estimates of trend.

In addition to facilitating reliable estimation of parameters, an effective monitoring program must also be efficient. Although too little sampling effort will allow only catastrophic trends to be detected, over-sampling wastes limited resources available to monitoring (Bernstein and Zalinski 1983). Therefore, sampling design represents a

balance between statistical credibility and fiscal reality (Committee on Environment and Natural Resources 1997). The extensive monitoring requirements that must be met under budgetary constraints has made cost-effective sampling design a research priority (Usher 1989, Committee on Environment and Natural Resources 1996, Schneider et al. 1999). Here, the efficiency of allocating effort to various sample effort components is evaluated to determine cost-effective sampling strategies.

Thesis Outline

This chapter has provided the context necessary for an exploration of the efficient design of large-scale avian monitoring programs. The importance of ecosystem monitoring has been established, the suitability of birds as monitoring targets has been outlined, and the influence of sampling design on the utility of monitoring programs has been stressed. In Chapter 2, a model for data collected from a large-scale avian monitoring program is developed, emphasizing sources of data variability. Empirical data sets from the forested region of Alberta are used to estimate parameters in the model for eight bird species and three avian community metrics. This facilitates estimation of the effectiveness of monitoring programs in Chapter 3. Statistical power is used to assess the effectiveness of monitoring programs, and Monte Carlo simulations are used to estimate statistical power for a range of candidate sampling designs. A model to estimate the cost of monitoring programs is developed in Chapter 4, and applied to candidate sampling designs. Power and cost estimates are then integrated, facilitating derivation of efficient sampling design strategies. An evaluation of the generality of these strategies is made possible by the inclusion of multiple bird species and community metrics in the analysis. In the final chapter, results from the previous three chapters are summarized and their implications for the design of large-scale avian monitoring programs discussed. The potential for expanding the analysis to additional taxa is evaluated, and an alternative method for determining efficient sampling strategies is introduced.

Chapter 2

Variability in Avian Abundance and Community Metrics

Introduction

An effective monitoring program must provide informative data. The amount of information obtained from a monitoring program depends upon sample size and variation in the data (Samways et al. 1995). In general, sample size is positively related, and variation negatively related, to data quality. Monitoring program design can thus be simplified to specifying the sample size that provides adequate data quality to achieve program goals, given the expected data variability.

Variability in monitoring data has two sources: natural variation and sample variation. Natural variation is due to spatial and temporal heterogeneity in the distribution of individuals. When monitoring is undertaken to detect changes in the abundance and/or distribution of species due to anthropogenic factors, natural variability may obscure these patterns. For example, consider a bird population that is decreasing over time due to habitat loss. Overlaid on this negative trajectory are short-term population fluctuations due to variable biotic and abiotic conditions (e.g. prey abundance and weather). If these short-term changes are large compared to the long-term trend, detection of the trend may be difficult. Sample variation occurs due to incomplete counts (Thompson et al. 1998). Incomplete counts must often be used as surrogates for actual population sizes, because census data are unobtainable, and this inaccuracy adds variability to data (e.g. Link et al. 1994). For example, repeated counts at a site will likely vary even if the actual number of individuals is consistent. This variability will, in general, decrease as the amount of effort spent obtaining a count increases.

A more precise description of expected variability in data requires specifying the structure of the monitoring program. The sampling design for a large-scale avian monitoring program can be generalized as follows (Figure 2.1). Multiple sites within the

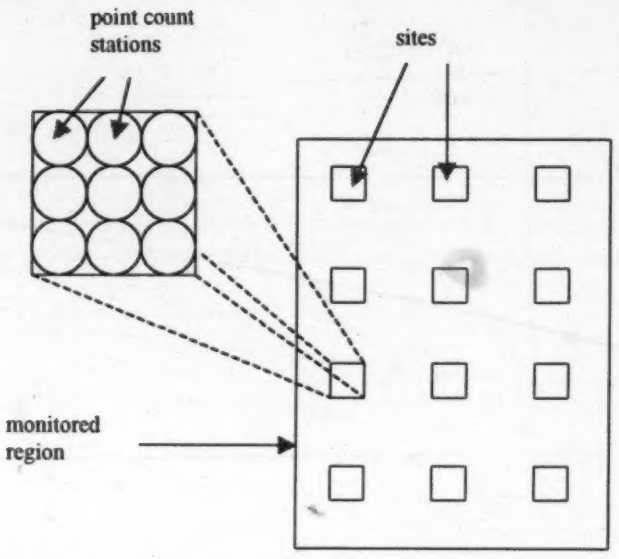

Figure 2.1. Generalized design of a large-scale avian monitoring program.

region of interest are sampled each year, or less frequently, during the breeding season (May and June). At each site, a certain level of sample effort is exerted in the form of multiple point count stations per site and multiple site revisits (surveys) within a sample year. Natural variation in counts across sites and through time will add noise that obscure trends. Incomplete counts at each site will also contribute variability. As effort exerted per site per year increases, either by sampling more stations or surveys, this variability should decrease. Under this sampling design framework, an understanding of the expected variability in monitoring data requires estimation of count variance across sites, count variance at a site across years, and the degree to which count variance decreases as sample effort increases.

In this chapter, variability in avian data collected from the specified sampling design is explored. Focus on the bird community is warranted because well-established monitoring protocols exist (e.g. Bibby et al. 1992) and bird species exhibit diverse and relatively well-understood responses to disturbance (Steele et al. 1984). Both population and community monitoring are considered. Population monitoring is valuable as it facilitates early detection of trends in biodiversity (Hagan 1992). Metrics that summarize

overall community structure, such as species richness or diversity measures, are also suitable for monitoring. Such metrics are less variable than populations and therefore require reduced sampling intensity (Karr and Chu 1997) and better communicate overall condition of the ecosystem (Karr 1981).

To explore variability in avian data, a model depicting expected variability in data from the sampling design is defined. Methods used to estimate the parameters in the model are then described. Using estimates of these parameters, eight bird species representing a range of variance characteristics are selected as targets for subsequent simulation studies. Finally, variance parameter estimates for these species, and for three community metrics, are presented and discussed.

The Model

Although there are a variety of purposes for monitoring programs, a common purpose is detection of long-term change, often due to anthropogenic factors. To design a monitoring program that fulfills this purpose, the expected variability about a population or community metric trajectory must be understood. In particular, simulation studies that explore the effectiveness of alternative sampling designs (Chapter 3) require accurate description of expected data.

Population trajectories often exhibit exponential growth or decline (Caughley and Sinclair 1994). Because growth or decline in a community metric reflects cumulative growth or decline in the populations that make up the community, community metric trajectories might also exhibit exponential growth or decline. I assumed that both population and community trajectories follow the exponential model:

$$Y_{it} = \alpha_0 e^{\beta t} + \varepsilon_{it} \quad \textbf{(2.1)}$$

where for the purposes described here,

Y_{it} = data collected at site i in year t,

α_0 = average abundance or community metric level at a site at the start of the monitoring period,

β = trend magnitude,

ε_{it} = error, $\varepsilon_{it} \sim N(0, \tau_1^2)$, and

τ_1^2 = within-site temporal variance.

The error term describes variability about the exponential trajectory. This variance, which I refer to as within-site temporal variance (τ_1^2), accounts for sample error and natural variation in abundance or the community metric across years. Sample error depends upon the accuracy of counts. Therefore, τ_1^2 is expected to decrease as effort exerted to collect data at each site increases.

The above model (Equation 2.1) assumes trajectories occurring at each site are identical. It is more likely that trajectories will differ both in initial levels prior to commencement of the trend and magnitude of the trend itself. These between-site variance components are due to differences in conditions across sites. The following model accounts for between-site variability:

$$Y_{it} = (\alpha_0 + \delta_i)e^{(\beta+\gamma_i)t} + \varepsilon_{it} \qquad \textbf{(2.2)}$$

where

δ_i = variability in initial abundance; $\delta_i \sim N(0, \tau_2^2)$,

γ_i = variability in effect size; $\gamma_i \sim N(0, \tau_3^2)$,

τ_2^2 = between-site variation in initial abundance or the community metric, and

τ_3^2 = between-site variation in trend magnitude.

Investigation of candidate sampling designs for monitoring thus require accurate estimation of τ_1^2, τ_2^2, τ_3^2, and α_0 and an evaluation of how τ_1^2 and α_0 are affected by sample effort.

Methods

Species and Community Metric Selection

Evaluating sampling design for all species within the bird community is logistically infeasible. Therefore, a subset of bird species was selected to represent a range of levels of within-site temporal and between-site spatial variance. This selection process was

completed using a plot of species' τ_1^2 and τ_2^2 estimates. Estimates of τ_1^2 used in the plot refer to a within-site effort of 20 point count stations per site and four surveys per year.

Species richness, richness of the ground-nesting guild, and the Shannon-Weiner diversity index were the community metrics selected for investigation. Species richness is often used in monitoring because it is easy to calculate and because it varies with many biotic and abiotic factors including productivity, habitat heterogeneity and complexity, and disturbance (Schluter and Ricklefs 1993). Overall species richness of a community, however, is an insensitive metric if disturbance alters community composition but not overall richness (e.g. Croonquist and Brooks 1991). Monitoring species richness of guilds allows community composition to be measured. Guilds are segments of the community that utilize resources in a similar way (Root 1969). Of five indices selected by Moors (1993) as potential avian indices of the integrity of lakes, four were based on guild structure, illustrating the utility of the guild concept to monitoring. Here, the ground-nesting guild was selected as an example. Species in this guild were the Black and White Warbler (*Mniotilta varia*), Blackpoll Warbler (*Dendroica striata*), Canada Warbler (*Wilsonia Canadensis*), Clay-colored Sparrow (*Spizella pallida*), Connecticut Warbler (*Oporornis agilis*), Dark-eyed Junco (*Junca hyemalis*), Hermit Thrush (*Catharus guttatus*), LeConte's Sparrow (*Ammodramus leconteii*), Lincoln's Sparrow (*Melospiza lincolnii*), Mourning Warbler (*Oporornis Philadelphia*), Northern Waterthrush (*Seiurus noveboracensis*), Orange-crowned Warbler (*Vermivora celata*), Ovenbird (*Seirus noveboracensis*), Palm Warbler (*Dendroica palmarum*), Ruffed Grouse (*Bonasa umbellus*), Song Sparrow (*Melospiza melodia*), Swamp Sparrow (*Melospiza georgiana*), Tennessee Warbler (*Vermivora peregrina*), Wilson's Warbler (*Wilsonia pusilla*), White-throated Sparrow (*Zonotrichia albicollis*), and Yellow-bellied Flycatcher (*Empidonax flaviventris*). Diversity indices, which combine species richness and the distribution of abundances across species into a single measure, are often preferred to richness metrics because distribution of abundance is often more sensitive to environmental stress than species richness alone (Magurran 1988). As an example of a diversity index, I selected the frequently used Shannon-Weiner function (Margalef 1958):

$$H = -\sum_{i=1}^{S} (p_i)(\log_2 p_i) \qquad \textbf{(2.3)}$$

where

H = Shannon-Weiner index of species diversity

S = number of species

p_i = proportion of total individuals in the community belonging to the i^{th} species

Estimating Parameters

τ_1^2, τ_2^2, τ_3^2, and α_0 reflect intrinsic properties of the system being monitored and should be estimated using data collected from the region of interest. Because I was interested in evaluating sampling design for monitoring programs in the boreal forest of Alberta, data collected from this region were used to derive the required parameter estimates. All variance parameters were estimated as coefficients of variation, to make them suitable for use in simulation studies. A coefficient of variation is calculated by dividing the standard deviation by the mean, resulting in a dimensionless measure of variability that is scaled relative to the mean.

Within-site Temporal Variation and Mean Initial Abundance

τ_1^2 and α_0 were estimated using data collected at three 100-ha control sites near Calling Lake, in north-central Alberta, over a six-year period (1993-1998). The sites, part of a long-term experimental fragmentation study (Schmiegelow et al. 1997), occur in boreal mixed-wood forest, an ecotype representative of much of northern Alberta. Control sites were used to minimize the effect of non-natural (treatment) variation, which can exaggerate within-site temporal variation (Gibbs et al. 1998). At each site, 20 point count stations were sampled four times (surveys) each year during the breeding season, at 10-day intervals from late May to late June. At each station, all birds seen or heard within a 100-m radius of the observer during a five-minute sampling interval were recorded. For additional details of the sampling design, see Schmiegelow et al. (1997).

To estimate α_0 and τ_1^2, I calculated the mean response and coefficient of variation at each site across the six-year period, and averaged these estimates across the three sites. The response at a site in a given year was total detections across all point count stations

averaged across surveys. For community metrics, this response was converted into species richness, richness of the ground-nesting guild, or the Shannon-Weiner function. If mean response at a site was zero across the six-year period, as was occasionally the case for rare species, the site was excluded from the analysis to avoid undefined coefficients of variation. To evaluate the effect of within-site sample effort on τ_1^2 and α_0, sub-samples were taken to reduce the number of stations and surveys.

To reduce the number of stations, sub-samples representing each possible contiguous arrangement of the reduced number of stations were taken from each site. To reduce the number of surveys, two strategies were used. To simulate a monitoring program where the timing of site visits can be kept constant from year to year, survey two was always sampled, and surveys three, four, and one were added successively as number of surveys increased from two to four. To simulate a monitoring program where timing of site visits cannot be kept constant from year to year, the surveys were randomly selected each year, without replacement. This process was repeated 100 times to achieve 100 sub-samples of randomly selected surveys across the six years. The sub-sampling procedure was performed for densities of four, nine, and 16 stations per site and one, two, three, and four surveys per year. The sub-samples were then used to estimate τ_1^2 and α_0. For the randomly scheduled scenarios, the mean estimates of τ_1^2 and α_0 across the 100 sub-samples were used as final estimates.

When sub-sampling to reduce the number of stations, multiple estimates of τ_1^2 and α_0 were generated for each site, to represent the various contiguous arrangements of the reduced number of stations. The mean of these estimates was taken as an overall estimate for each site. Any configuration for which no individuals were counted over the six-year period was removed from the analysis to avoid undefined coefficients of variation.

Between-site Variation

Between-site variation in initial abundance (τ_2^2) and between-site variation in trend (τ_3^2) were estimated using the Breeding Bird Survey (BBS) database (USGS Patuxent Wildlife Research Center 2000). The BBS is a volunteer based long-term monitoring

program spanning Canada and the United States, designed to provide information on the distribution and abundance of breeding birds (Droege 1990). The program consists of routes that are sampled by volunteers each year during the breeding season (May or June). At each route, birds are sampled at 50 point count stations, spaced one-half mile apart, along 24.5-mile routes. More information on the BBS design is provided by Droege (1990). BBS routes in the closed boreal forest physiographic region of Alberta (USGS Patuxent Wildlife Research Center 2000) were used. The identification numbers of routes included in the analyses are 32, 33, 34, 42, 49, 50, 62, 63, 77, 78, 122, 131, 141, 231, 241, and 251.

τ_2^2 was estimated as response variance between routes within a year. This variance component reflects not only variability between routes, but also variability due to sampling error at each route (Link et al. 1994). Because sampling error at each site is already accounted for by within-site variation, I isolated the spatial variability in abundance by multiplying variance estimates by (1 - proportion of between route variance due to sampling error). The sampling error associated with the BBS design is likely larger than that associated with the sampling design considered here (Figure 2.1), so this had the added benefit of reducing the implications of differences between the two sampling designs. Estimates of the proportion of between BBS route variance due to sampling error (λ) were taken from Link et al. (1994), who estimated this quantity using repeated counts at BBS routes within seasons. For species not evaluated by Link et al. (1994), λ was estimated using Link et al.'s (1994) finding that λ is negatively related with abundance and their suggested model for this relationship, a regression of logit(λ) on the natural logarithm of abundance. Using estimates of λ and average counts per species per route presented in Link et al. (1994), I performed the regression, which resulted in the following model:

$$y = 0.4827 - 0.7420x \qquad \textbf{(2.4)}$$

where y equals logit(λ) and x equals the natural logarithm of abundance. For species not included in Link et al.'s (1994) study, Equation 2.4 was used to estimate logit(λ), letting x equal the natural logarithm of the average count per route visit for a species. Estimates of λ were achieved by transforming logit(λ). Sampling error could not be extracted for the community metrics because Link et al. (1994) did not evaluate sampling error at the

community level. After extracting sampling error, the coefficient of variation in response across routes was calculated. Estimates were made using data from 1992 to 1998 (11-16 routes per year), and the average of these coefficients of variation was used to estimate τ_2^2. To avoid undefined coefficients of variation, years with zero mean response across routes were excluded from the analysis.

τ_3^2 was estimated using variation in trend across BBS routes in the closed boreal forest region of Alberta between 1989 and 1998. Although counts from 1989 to 1991 were not used to estimate τ_2^2 due to reduced numbers of routes sampled in this time frame, they were used here to achieve longer trajectories to enhance τ_3^2 estimates. No counts before 1989 were used due to the infrequency of route visits in the region during years prior to 1989. For species, only routes satisfying the United States National Biological Service data selection criteria (Thomas 1997a) were used in the analysis. In addition to restricting analysis to routes within a specified range of dates, times and weather conditions, the criteria ensure adequate data have been collected to permit analysis (i.e. number of surveys minus number of observers must be greater than one and at least one individual of the species of interest must have been observed at the route). After selection criteria had been applied, data from four to 15 routes were available, depending on the species. All 15 routes satisfying the date, time, weather condition, and count intensity criteria were used for the community metrics. The trend at each route was estimated by fitting the data to the exponential model. A computer program written by Thomas (1997b) was used for the analysis. After estimating the variance of route trend, a coefficient of variation was calculated by dividing the square root of the variance by the mean route trend.

Distribution Assumptions

Equation 2.2 assumes ϵ_{it}, δ_i, and γ_i are normally distributed random variables. A normal distribution was assumed for γ_i because histograms of route trend estimates for species and community metrics displayed approximate bell-shaped curves. Normal distributions, truncated at zero, were assumed for ϵ_{it} and δ_i. Although discrete distributions, such as the Poisson, are often used for count data, a continuous distribution

was required because continuous data result when counts are averaged across multiple surveys prior to analysis. For species, the normal and lognormal distributions were considered. The normal was selected, however, because cumulative distribution functions for the distribution of species over time and across sites suggested the lognormal was inappropriate due to under representation of zero counts. For community metrics, the normal distribution was selected on theoretical grounds. All three metrics are calculated by summing random variables. According to the central limit theorem, statistics calculated by summing random variables are distributed normally (Hilborn and Mangel 1997).

Results

Species Selection

The species variance plot (Figure 2.2) was divided into quadrants by the median of τ_1^2 estimates and the mean of τ_2^2 estimates across all species, except those with maximum τ_1^2 (2.449). These species were assumed inappropriate targets for large-scale monitoring due to their rarity (observed during a maximum of one of the six years at each site). The excluded species were the Alder Flycatcher (*Empidonax alnorum*), Boreal Chickadee (*Poecile hudsonicus*), Common Yellowthroat (*Geothlypis trichas*), Eastern Phoebe (*Sayornis phoebe*), Olive-sided Flycatcher (*Contopus cooperi*), Orange-crowned Warbler (*Vermivora celata*), Three-toed Woodpecker (*Picoides tridactylus*), White-winged Crossbill (*Loxia leucopiera*), and Yellow-bellied Flycatcher (*Empidonax flaviventris*). The median of τ_1^2 was used instead of the mean due to the skewed distribution of this variance parameter across species. The plot was interpreted as follows: species in the lower left hand quadrant are ubiquitous, exhibiting low spatial and temporal variance; species in the lower right hand quadrant fluctuate over time, exhibiting high temporal variance and low spatial variance; species in the upper left hand quadrant have patchy distributions, exhibiting high spatial variance and low temporal variance; and, species in the upper right hand quadrant are irruptive, exhibiting high temporal and spatial variance.

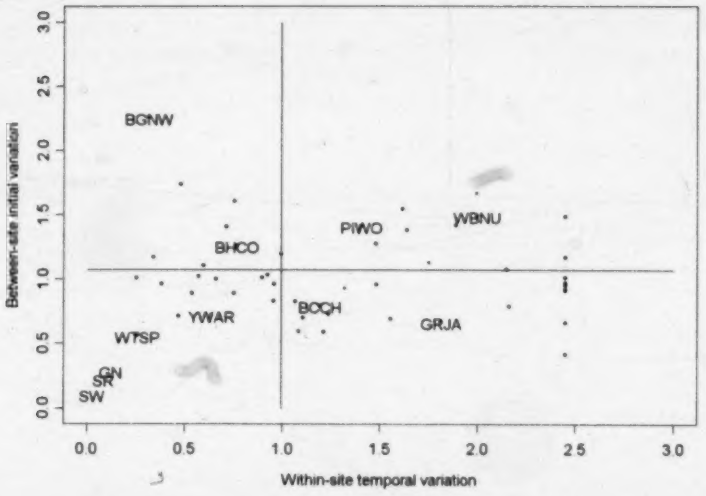

Figure 2.2. Temporal and spatial variability in bird species abundance and community metrics. The horizontal and vertical lines dissecting the plot represent mean between-site initial variation and median within-site temporal variation, respectively, across all species, excluding those with maximum within-site temporal variation (2.449). The four letter codes identify the eight species and three community metrics selected as targets for simulation studies (BCCH = Black-capped Chickadee, BGNW = Black-throated Green Warbler, BHCO = Brown-headed Cowbird, GRJA = Gray Jay, PIWO = Pileated Woodpecker, WBNU = White-breasted Nuthatch, WTSP = White-throated Sparrow, YWAR = Yellow Warbler, SR = species richness, GN = richness of the ground-nesting guild, SW = Shannon-Weiner index).

The White-throated Sparrow (*Zonotichia albicollis*), Gray Jay (*Perisoreus canadensis*), Black-throated Green Warbler (*Dendroica virens*), and White-breasted Nuthatch (*Sitta carolinensis*) were selected as extreme examples of ubiquitous, temporally fluctuating, spatially patchy, and irruptive species respectively. The Yellow Warbler (*Dendroica petechia*), Black-capped Chickadee (*Parus atricapillus*), Brown-headed Cowbird (*Molothrus ater*), and Pileated Woodpecker (*Dryocopus pileatus*) were selected as less extreme examples of ubiquitous, temporally fluctuating, spatially patchy, and irruptive species, respectively. The range of species outlined above allows comparisons of sampling requirements for species exhibiting different variance levels and for rare versus common species. In addition to the selected species, three community metrics were included in the analysis. The variability of species richness, richness of the

ground-nesting guild, and the Shannon-Weiner diversity index was lower than the variability of species abundances (Figure 2.2).

Parameter Estimates

Estimates of τ_1^2, τ_2^2, τ_3^2 and α_0 for the selected species and community metrics are provided in Appendix 1. To summarize the effect of within-site sample effort on τ_1^2, the response of τ_1^2 to number of surveys, number of stations, and timing of surveys was calculated for each species (Table 2.1) and community metric (Table 2.2). The response of τ_1^2 to number of surveys was evaluated as the percent reduction in τ_1^2 caused by increasing number of surveys from one to four, averaged across the three station levels. This was done for both fixed and variable survey scheduling scenarios to test the hypothesis that number of surveys is more important when timing of data collection is not consistent. The response of τ_1^2 to number of stations was evaluated as the percent reduction in τ_1^2 caused by increasing the number of stations from four to 16, averaged across the four survey levels. To evaluate the influence of timing of surveys, the percent reduction in τ_1^2 caused by using fixed rather than randomly scheduled surveys, averaged across all effort levels except when surveys equaled four, was calculated.

Species	% reduction in τ_1^2 due to surveys when timing of surveys is fixed	% reduction in τ_1^2 due to surveys when timing of surveys is random	% reduction in τ_1^2 due to stations when timing of surveys is random	% reduction in τ_1^2 due to timing of surveys
BCCH	28.79%	24.96%	27.44%	-4.53%
BGNW	46.46%	47.24%	45.56%	-5.39%
BHCO	26.17%	28.11%	33.65%	0.44%
GRJA	-22.37%	17.31%	15.23%	21.01%
PIWO	28.75%	26.31%	21.58%	-13.88%
WBNU	15.31%	9.50%	6.66%	1.67%
WTSP	45.05%	39.36%	24.59%	-5.46%
YWAR	44.70%	37.43%	28.59%	-12.39%
AVERAGE	26.61%	23.33%	25.41%	-0.02%

Table 2.1. Percent reductions in τ_1^2 of species due to increasing surveys from one to four when timing of surveys are fixed, increasing surveys from one to four when timing of surveys are variable, increasing stations from four to sixteen when timing of surveys are variable, and using fixed rather than randomly scheduled surveys. Species codes as in Figure 2.2.

Community metric	% reduction in τ_1^2 due to surveys when timing of surveys is fixed	% reduction in τ_1^2 due to surveys when timing of surveys is random	% reduction in τ_1^2 due to stations when timing of surveys is random	% reduction in τ_1^2 due to timing of surveys
SR	30.93%	37.55%	25.72%	17.20%
GN	42.93%	29.55%	26.24%	4.07%
SW	48.49%	53.52%	43.09%	21.29%
AVERAGE	40.78%	40.21%	31.68%	14.19%

Table 2.2. Percent reductions in τ_1^2 of community metrics due to: increasing surveys from one to four when timing of surveys are fixed, increasing surveys from one to four when timing of surveys are variable, increasing stations from four to sixteen when timing of surveys are variable, and using fixed rather than randomly scheduled surveys. Community metric codes as in Figure 2.2.

Discussion

Summary of Results and Implications for Monitoring

Averaged across both the species and community metrics analyzed, the between-site initial variance component was greatest in magnitude, followed by the within-site temporal variance component and the between-site trend variance component (Table 2.3). Within-site temporal variance was estimated at multiple within-site effort levels and requires further discussion.

	τ_1^2	τ_2^2	τ_3^2
Across species	1.083 (0.642)	1.143 (0.575)	0.303 (0.133)
Across community metrics	0.0787 (0.0494)	0.203 (0.0910)	0.0507 (0.0130)

Table 2.3. Mean (and standard deviation) estimates of τ_1^2, τ_2^2, and τ_3^2 across selected species and community metrics.

The Influence of Within-site Effort on Within-site Temporal Variance

To evaluate the influence of timing of surveys, number of surveys, and number of point count stations on within-site temporal variance (τ_1^2), this variance component was estimated across a range of within-site effort levels. I expected timing of surveys to

affect τ_1^2 in two ways. First, variable timing of surveys was expected to inflate τ_1^2 because the species chronology of the sampled community would differ from year to year. This pattern occurred for only three species: Brown-headed Cowbird, Gray Jay, and White-breasted Nuthatch. On average, across all species, variable timing of surveys actually decreased τ_1^2 by 0.02%. The expected pattern was evident for community metrics, however, with τ_1^2 increasing, on average, by 14.19%. Second, τ_1^2 was expected to be more sensitive to the number of surveys sampled when timing of surveys was variable, because increasing number of surveys would decrease the assumed inflation of τ_1^2 described above. The expected pattern was not evident for species or community metrics. On average, τ_1^2 exhibited similar sensitivity to number of surveys regardless of whether surveys were fixed or random (Tables 2.1 and 2.2). The insensitivity of τ_1^2 to timing of surveys for species suggests within-breeding season patterns in bird abundance are sufficiently variable from year to year to reduce the importance of a consistent monitoring schedule. This result is fortunate for the design of large-scale avian population monitoring programs, where consistent timing of surveys is difficult to achieve. However, if possible sample dates extended beyond the one month period during which data were collected at the Calling Lake research station, the importance of surveys would almost certainly increase. The increased sensitivity of τ_1^2 to timing of surveys for community metrics suggests within-season patterns in overall structure of the community are more consistent across years. Despite the importance of timing of surveys to monitoring community metrics, scheduling of surveys is assumed random for the remainder of the discussion. Achieving consistent timing of surveys from year to year is logistically difficult, so randomly timed surveys were considered more realistic.

As expected, τ_1^2 decreased as number of surveys and number of stations increased. Averaged across species, increasing the number of surveys from one to four caused a 23.33% decrease in τ_1^2, while increasing the number of stations from four to 16 caused a 25.41% decrease in τ_1^2. This suggests the improvement in data quality achieved by increasing number of surveys from one to four and increasing number of stations from four to 16 is very similar, which may have important implications for allocation of sample effort because of the greater expense of increasing surveys as compared to stations. τ_1^2 of the community metrics also decreased with increased within-site effort.

However, the decrease due to number of surveys (40.21%) was greater than the decrease due to stations (31.68%).

The Variability of Species

Using a plot of within-site temporal and between-site spatial variance, species representing four variance categories were selected. For the ubiquitous category, the White-throated Sparrow and Yellow Warbler were selected. This category likely represents generalist species whose broad niches facilitate occupancy of a wide range of habitats. For example, the White-throated Sparrow is an abundant species in a variety of habitats. The Black-throated Green Warbler and Brown-headed Cowbird were selected as examples of spatially patchy species. Species in this category are likely spatially variable because their narrow niches allow occupancy of only a subset of available habitat. Occupancy within this subset of habitats, however, is consistent across years. For example, the Brown-headed Cowbird is highly successful provided the habitat is not densely forested (Lowther 1993). As examples of temporally fluctuating species, the Black-capped Chickadee and Gray Jay were selected. These species may depend upon resources that vary temporally over large spatial scales. This category may also contain generalist species associated with high sample error, causing inflation of within-site temporal variance. To represent irruptive species, the Pileated Woodpecker and White-breasted Nuthatch were selected. The quality of habitat for these species may be variable, such as is the case with cone crops and insect outbreaks. In addition, sampling issues may contribute to the high variance of species belonging to this category. For example, the Pileated Woodpecker is not an irruptive species per say, but its large home range relative to the scale of point count stations results in variable counts, even in the presence of a stable population. In addition, species in this category are rare, causing them to be associated with high sample error, which inflates within-site temporal variance. Unfortunately, the species whose ecology makes them of greatest monitoring interest are also species for which variance is highest and therefore sampling requirements will be greatest. Generalist species, such as the White-throated Sparrow, will be easiest to monitor. Their broad habitat requirements, however, may also imply

insensitivity to disturbance, making these species poorly suited for monitoring. Specialist species, such as the Pileated Woodpecker, will require greater sampling intensity to mitigate high natural variability and sample error.

The eight species selected also facilitate a comparison of variability for rare and common species. To explore these relationships, species were ranked in order of abundance and the three variance parameters. The abundance ranking was then compared with the ranking for each variance parameter by calculating a correlation coefficient. There was a strong negative relationship between abundance and within-site temporal variation rankings ($r = -0.976$ ($p = 0.000$)). Ranking of abundance and the other two variance parameters were also negatively correlated, although the relationships were weaker ($r = -0.714$ ($p = 0.047$) for between-site trend variance and $r = -0.476$ ($p = 0.233$) for between-site initial abundance variance). These relationships indicate that, in general, less abundant species are more variable. Two possible reasons for this pattern are: rare species are often specialists, making them susceptible to variability in environmental conditions; and rare species are associated with high sample error (Link et al. 1994), which inflates within-site temporal variability. In addition to being associated with high variability, rare species are usually of conservation interest. Once again, species requiring the highest sample effort will often be species that require monitoring.

The Variability of Community Metrics

The high variability of populations as compared to community metrics has led to claims that population monitoring is an ineffective way to measure anthropogenic disturbance to ecosystems (e.g. Karr and Chu 1997). The results presented here support this view. However, two factors may limit the utility of community metrics. First, the same insensitivity to variable conditions that reduces the natural variability of community metrics also reduces their sensitivity to disturbance. For example, a much larger disturbance is required to reduce species richness by 10% than a population by the same amount. This insensitivity to disturbance may nullify the advantage of low natural variation and sample error. Second, the response of community metrics to disturbance is not well understood. Diversity metrics, such as the Shannon-Weiner function, are

thought by some to be difficult to interpret because they confound two community variables, evenness and richness (e.g. Karr and Chu 1997). Species richness, on the other hand, does not consider the abundance of species, which is often more sensitive to disturbance than richness (Magurran 1988). Despite these limitations, the low noise associated with community metrics suggests further research into the use of community metrics to monitor ecosystem response is warranted.

Methodology

Evaluation of sampling design effectiveness is sensitive to variance parameter estimates (Chapter 3). Therefore it is important to accurately estimate these parameters.

Within-site Temporal Variance

Here, I estimated within-site temporal variance by calculating variance in abundance across years at three control sites. Control sites, which were assumed devoid of population trends, were used to eliminate the confounding effects of trend on variance. A check of the assumption supported its validity, with 95% significant trends being absent for the eight species selected. Instead of this method, Gibbs et al. (1998) estimated temporal variance from the residuals of a linear regression of counts against time. The theory behind this strategy is that it removes trend from the counts and therefore minimizes the effect of non-natural variation. However, estimating variance from the residuals of a linear regression is only valid if a significant trend exists. Otherwise, the method can underestimate temporal variance by allocating natural variation to trend that does not exist. This may explain why Gibbs et al.'s (1998) estimate of temporal variation for small birds (coefficient of variation = 0.57) was lower than estimates presented here (mean coefficient of variation across species = 1.083). Underestimating temporal variance results in overly optimistic predictions of the effectiveness of proposed monitoring strategies. I feel temporal variance as estimated from control sites provides more accurate estimates of the parameter and results in more conservative assessments of proposed monitoring programs.

Within-site temporal variance incorporates sampling error and natural variation in actual number of birds over time. By sub-sampling an empirical data set, I was able to evaluate how number of stations per site and number of surveys influenced sampling error (and therefore within-site temporal variance). The large decrease in temporal variance achieved by increasing within-site effort indicates that attention to within-site effort levels was warranted. Additional components of within-site effort can also influence measurement error and require consideration. Most important are sampling technique and length of point counts. The variance estimates reported here assume point counts are used to collect data. Point counts are the preferred method for large-scale monitoring in forested regions (Verner 1988). However, spot mapping and transects are two additional techniques commonly used to count birds (Bibby et al. 1992). Spot mapping is inappropriate due to the intensity of labor required (Verner 1988, Bibby et al. 1992). Advantages of point counts over transects include improved detection of birds, improved spatial coverage of sites, and increased compatibility with habitat data (Bibby et al. 1992). Due to the widespread use of point counts and their suitability for large-scale avian monitoring, the dependence of estimates on point count methodology should not restrict generality of these results.

The estimates of within-site temporal variance also assume a sampling interval of five minutes at each point count station. Using forest birds in Hawaii, Barker et al. (1993) determined the optimal sampling time per station to be 4.663 minutes if the goal of the study was to detect population change. In addition, Dawson (1981) reported little difference in sampling efficiency between counts of two, five, and ten-minute lengths. Due to the suitability of five-minute sampling periods for monitoring population change, and the minimal influence of point count duration on efficiency, the use of five-minute point counts in this analysis should also not restrict generality.

Between-site Initial Variance

Between-site variation in initial abundance (τ_2^2) estimates the natural variation in abundance across sites at the start of a monitoring period. I used data from BBS routes to estimate this variance component, which is problematic because between-route spatial

variance is confounded with within-route sample error, causing an inflation of τ_2^2. For species, this problem was avoided by using Link et al.'s (1994) estimates of the proportion of between-route spatial variance due to within-route sample error. Although estimates of this proportion are not provided for all species encountered on the BBS routes I analyzed, estimates were provided for all selected species, except the Gray Jay. In this case, an estimate of between route variance due to sampling error based on a relationship with abundance (Equation 2.4) was used. Within-route sample error could not be extracted for community metrics because they were not part of Link et al.'s (1994) study. Therefore, between-site variation in initial abundance estimates for community metrics may be exaggerated.

Between-site Trend Variance

Estimation of between-site variation in trend is difficult due to data requirements. Data are required from multiple sites (three or more) for the same three or more years. No empirically based estimates of between-site variation in trend could be found in the literature, although Gibbs et al. (1998) provided an unsubstantiated estimate of coefficient of variation equal to 0.10. I used the BBS database to estimate the parameter because it is the best, and likely the only, data source available. This resulted in an average between-site variation in trend (expressed as a coefficient of variation) of 0.303 across species and 0.0507 across community metrics.

Data Limitations

The best available data were used to make the variance estimates presented in this chapter. Data limitations, however, required use of the BBS database. BBS sampling design is not identical to the generalized sampling design assumed here and some inaccuracy in the estimates may be associated with this difference. In addition, sites with different cover types than the sites at Calling Lake may display different temporal variance. Improved data would improve these deficiencies. The most efficient way to improve the data would be to implement a long-term monitoring program whose design

was based on the estimates presented here. After a suitable time period had elapsed (i.e. five years), data from the monitoring programs could be used to re-estimate variance parameters. These estimates could then be used to reassess monitoring program design, and the necessary changes could be implemented.

Variance estimation was focused on the bird community, in part because of the suitability of birds to monitoring, but also due to data availability. Data required to estimate the variance parameters (Table 2.4) are in short supply for taxa other than birds. Data are usually collected from sites for only one or two years, and are therefore inappropriate for estimation of τ_1^2 and τ_3^2. Although data are often collected from multiple sites, the spatial separation of the sites is usually too small to facilitate estimation of τ_2^2 or τ_3^2. Spatial separation should be at least the minimum distance between sites for the proposed monitoring program, so that true spatial variation in flora and fauna is accounted for. The only likely source of data with the required spatial and temporal extent is monitoring programs. However, long-term monitoring programs are rare. It may be necessary, therefore, to use theoretically based variance parameter estimates for other taxa when assessing monitoring program design. As suggested above, data from the implemented monitoring program could then be used to estimate the variance parameters after a certain time period, and monitoring program design could be adjusted accordingly.

Variance parameter	Minimum number of sites	Minimum number of years
τ_1^2	1	3
τ_2^2	3	1
τ_3^2	3	3

Table 2.4. Spatial and temporal extent of data sets required to estimate τ_1^2, τ_2^2, and τ_3^2.

Summary

A model for data collected from a large-scale avian monitoring program was developed and empirically based estimates of the model's parameters were presented for eight species, representing a range of variance characteristics, and for three community metrics. The estimates indicate that species most likely to be monitored, such as rare

species and habitat specialists, exhibit higher natural variability and therefore will require increased sample effort to achieve high data quality. Community metrics, on the other hand, exhibited low natural variability and may require lower sample effort, although their value is questionable due to the insensitivity of these metrics to disturbance. For species and community metrics, sample error decreased substantially as number of surveys and number of point count stations increased, emphasizing the importance of these components of within-site sample effort. In contrast, the ability to consistently time surveys across years did not consistently decrease sample error. Despite limitations due to data availability, the variance estimates made the best use of existing data and should facilitate reasonably accurate simulation of data collected from an avian monitoring program in the boreal forest of Alberta. This, in turn, will allow an assessment of avian monitoring program design for this region.

Chapter 3

The Effect of Sample Effort on Power

Introduction

Anthropogenic disturbance of ecosystems and insufficient understanding of the effects of this disturbance (Walters and Holling 1990) necessitates ecological monitoring. Monitoring how ecosystems respond to disturbance provides information vital for assessing the impact of management actions and for identifying elements of biodiversity at risk. This facilitates adaptive management and the setting of conservation and research priorities.

Recognizing the important roles of ecological monitoring data, policy makers have called for increased monitoring of ecosystems in documents such as the Convention on Biodiversity (Johnson 1993) and the Canadian Biodiversity Strategy (Minister of Supply and Services 1995). Despite these commitments, large-scale ecological monitoring programs are rare (Schneider et al. 1999) and no adequate program currently exists to effectively monitor Alberta's forested regions (Lee and Hanus 1998). An important limiting factor is the challenge in identifying sampling designs that provide useful data. This difficulty is highlighted by the Environmental Monitoring and Assessment Program (United States Environmental Protection Agency 2000), which was implemented to monitor status and trends of USA's ecological resources. Despite a large investment of effort and money into the program, a review by the National Research Council (1995) concluded the program utilized insufficient sampling intensity to achieve program goals.

To construct an effective sampling strategy, sample effort parameters must be related to monitoring effectiveness. This requires a measure to communicate the utility of data to monitoring. If the purpose of monitoring is detection of trend, effectiveness of an avian monitoring program can be defined as the probability of detecting a change in population size or community structure given a change has occurred. This is equivalent

to statistical power, where power is the probability of rejecting the null hypothesis (no trend) when the null hypothesis is false.

Two implications of ignoring power highlight its importance. First, if a low-power monitoring program is implemented, there will be little chance of finding statistically significant changes even when true changes are large (Peterman 1990, Taylor and Gerrodette 1993). This illusion of stability can prevent the remedial action required to arrest degradation of biodiversity (Fairweather 1991). Alternatively, ignoring power may lead to excessive sampling, thereby wasting effort. The importance of power suggests *a priori* power analysis is an effective tool for designing a long-term monitoring program (Peterman 1990, Zielinski and Stauffer 1996, Steidl et al. 1997, Gibbs and Melvin 1997, Gibbs et al. 1998).

Power depends on significance, effect size, and data variability (Cohen 1988). Significance is the probability of accepting the null hypothesis when it is true. Like power, significance is a desirable property. However, significance and power are negatively related, and represent an important tradeoff to consider when designing a monitoring program. Effect size is the minimum change in biodiversity to be detected. As effect size increases, the probability of detecting the effect increases. Data variability is negatively related with power. Biodiversity varies temporally and spatially within regions of interest, and this natural variation decreases power by obscuring patterns. Patterns are also obscured by sampling variation. Sampling variation occurs due to sampling only a portion of the sampling frame and incomplete counts at each site (Thompson et al. 1998). Sampling variation generally decreases as sample effort increases, resulting in a positive relationship between sample effort and power.

This chapter evaluates the design of large-scale avian monitoring programs. The sample effort components comprising the sampling design for a large-scale avian monitoring program were described in Chapter 2. Here, results from Monte Carlo simulations based on empirically derived parameter estimates are used to develop relationships between sample effort components and power to detect trends in populations of eight bird species and three bird community metrics. This facilitates an analysis of how sampling requirements differ across species with differing variance characteristics, and community metrics, and of the rate at which power increases as effort

is allocated to number of sites, sampling frequency, number of point count stations, and number of surveys.

Methods

Estimation of relationships between sample effort components and power to detect trend involved three steps. Simulations were conducted to estimate the power of alternative sampling designs. Multiple regression was then used to model relationships between power and sample effort components. Finally, partial derivatives of these relationships with respect to each sample effort component were calculated to estimate the rate at which power changed with respect to each sample effort component.

Monte Carlo Simulations

Sample effort was related to power using a Monte Carlo simulation approach. The Monte Carlo approach uses random sampling to solve problems involving processes affected by random factors (Sobol 1975). Here, the problem was estimating power of sampling designs, a quantity affected by random fluctuations in data due to natural variation and sampling error. Solution using the Monte Carlo approach required simulation of data collected from monitoring a population or community experiencing a trend. This was done by randomly sampling distributions describing monitoring data, following the strategy used by Gibbs et al. (1998). Simulated data were then analyzed in an attempt to detect the trend. This process was repeated numerous times, and power estimated as the proportion of simulated trajectories for which trend was correctly identified.

Assumed Data Structure

Monitoring occurred at multiple sites over multiple years, following the sampling design outlined in Figure 2.1. Simulated data trajectories followed the exponential model (see Chapter 2). Variability entered the monitoring data via three components. First,

data fluctuated at a site across years due to variability in conditions and sampling error. This component, termed within-site temporal variance (τ_1^2), was estimated at multiple within-site effort levels to discern how sample error changed with increased effort. Data fluctuated spatially in two ways. The distribution of individuals across sites at the start of the monitoring period was heterogeneous. Between-site initial variance (τ_2^2) accounted for this variability. Trend magnitude also varied across sites about the regional trend magnitude. This was accounted for by between-site trend variance (τ_3^2). The distributions of species or community metrics across years and sites, as well as the distribution of trends across sites, were assumed gaussian (Chapter 2). Assumed normal distributions for species and community metrics across years and sites were truncated at zero to avoid negative values. The methods used to estimate τ_1^2, τ_2^2, and τ_3^2 are summarized in Table 3.1. A more detailed description and parameter estimates are provided in Chapter 2.

Variance Parameter	Description	Data	Estimation Method
Within-site temporal variance (τ_1^2)	Variability in counts at a site over time due to natural variation and sample error.	Three 100 ha control sites at the Calling Lake Fragmentation Study (1993-1998)	• τ_1^2 = variance in counts at a site across years • estimated using full dataset and sub-samples to evaluate response of τ_1^2 to sampling intensity
Between-site initial variance (τ_2^2)	Natural variability in counts across sites at the start of the monitoring program.	BBS routes in the closed boreal physiographic region within Alberta (1992-1998).	• τ_2^2 = variance in counts across routes within a year • within-route variance was extracted using estimates provided by Link et al. (1994)
Between-site trend variance (τ_3^2)	Natural variability in trend magnitude across sites.	BBS routes in the closed boreal physiographic region within Alberta (1989-1998), subject to USNBS route selection criteria	• τ_3^2 = variance in trend across routes

Table 3.1. Description and estimation methods for within-site temporal variance, between-site initial variance, and between-site trend variance.

Simulation Process

Data simulation required specification of regional trend (β), initial abundance (α_0), variance (τ_1^2, τ_2^2, and τ_3^2), monitoring program length (I), number of sites sampled, and how frequently they were sampled. Although explicit specification of effort per site was not required, number of point count stations and surveys was implicit in τ_1^2 and α_0. This is because estimates of these parameters were made for multiple within-site effort levels. All variance parameters were entered as coefficients of variation, allowing variance to be scaled during simulations, as described below.

Individual trajectories were simulated for each site. Initial abundance at the site was determined by drawing a random number from a normal distribution, truncated at zero, with mean α_0 and standard deviation equal to $\alpha_0 * \tau_2^2$. Trend magnitude was similarly generated by drawing a random number from a normal distribution with mean β and standard deviation equal to $\beta * \tau_3^2$. The site's trajectory was then simulated across I years using the exponential model. Trajectories were truncated at zero. Using the trajectory, the datum collected from a site was generated for each year specified by the sampling design by drawing a random number from a normal distribution, truncated at zero, with mean equal to the abundance in the trajectory at that time step and standard deviation equal to abundance$*\tau_1^2$. This added temporal variance to the population trajectory. This process was completed for each site in the sampling design. If sampling occurred less frequently than every year, all sites were sampled in common years, and none in the intervening years.

To analyze the simulated data, route regression (Geissler and Sauer 1990) was used. Route regression is a two-stage estimator, whereby trends are first estimated at each site, and then combined to estimate regional trend. Trend at each site is estimated by fitting the data to the exponential model using least squares. Site trends were used to estimate regional trend using the same method as the United States National Biological Service when analyzing Breeding Bird Survey data, as described in Thomas (1997a). Regional trend was estimated as a weighted mean trend across sites, weighting site trends by their marginal mean count and precision. To estimate significance of regional trend, 400 bootstrap samples of the site trends were taken. A bootstrap sample is achieved by

repetitively sampling the original data with replacement. This provides an estimate of the sample distribution of the statistic, based on the assumption that the bootstrap samples are a suitable indication of the true distribution of the statistic (Thomas 1997a). For each bootstrap sample, the regional mean trend was calculated. Mean and variance in regional trend estimates across the bootstrap samples were then used to evaluate trend significance using a two-sided z-test to test the alternate hypothesis that regional trend was different than zero.

To estimate power of a sampling design, 1000 repetitions of the simulation process were completed (Link and Hatfield 1990). Power equaled the proportion of simulations for which trend was determined significant (p-value less than or equal to alpha) and trend direction (increasing or decreasing) correctly identified. Alpha was set at 0.20, representing an 80% significance level. This follows Gibbs et al.'s (1998) recommendation that power and significance be set at 90% and 80% respectively for monitoring programs. Traditionally, power and significance have been set at 80 and 95% respectively (e.g. Zielinski and Stauffer 1996, Hayes and Steidl 1996). It is important, however, to evaluate the relative costs of the two error types to ensure reasonable weighting of their importance (Peterman 1990). In biodiversity related issues, Type II errors are likely more damaging from a management perspective due to high recovery costs if remedial action is delayed and the high intrinsic cost of lost habitats or species (Fairweather 1991). In comparison, the cost of wasted research effort and unnecessary changes in management practices due to occasional false alarms caused by Type I errors may be minimal.

Monte Carlo simulations were performed using a program written in the C programming language. Code for the power analysis program is available upon request. Program code for the route regression method was provided by Thomas (1997b) and numerical recipes for random number generation (Press et al. 1992) were used.

Simulation Scenarios

Using parameter estimates presented in Chapter 2, simulations investigated power to detect trends in populations of eight bird species, as well as total bird species richness,

species richness of the ground-nesting guild, and the Shannon-Weiner diversity index. The species, selected to provide a range of variability characteristics (see Chapter 2), were the Black-capped Chickadee, Black-throated Green Warbler, Brown-headed Cowbird, Gray Jay, Pileated Woodpecker, White-breasted Nuthatch, White-throated Sparrow, and Yellow Warbler. For each species or community metric, 460 sampling designs were investigated (Table 3.2). Programs 10 and 20 years in length were simulated. Here, program length refers to the number of years the program spans so that a 10 year program actually has a maximum of 11 years of sampling. Because of the real logistic difficulties of sampling all sites during the same survey period, it was assumed that timing of monitoring varied randomly from year to year within the two-month sampling period (May and June).

Effort Parameter	Effort levels	
	10 year programs	20 year programs
Number of sites	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50
Sampling frequency (years)	1, 2, 3, 5	1, 2, 3, 4, 5
Number of point count stations	16	4, 9, 16
Number of surveys	1, 2, 3, 4	1, 2, 3, 4

Table 3.2. Sampling designs evaluated in power simulations.

For programs 20 years in length, number of sites varied between 10 and 50 in increments of 10, number of point count stations per site varied between 4, 9, and 16, and number of surveys varied between one and four, in increments of one. A minimum of 10 sites was used because this is the minimum number of sites recommended when using route regression (Geissler and Sauer 1990). Frequency of sampling varied between every year and every five years in increments of one year. However, sampling every three years results in sampling ending in the 18th year of the program. To avoid the negative bias this would cause due to sampling not covering the entire program length, sampling occurred on the 20th year for programs with sampling every three years, even though sampling should not have occurred until the 21st year. After assuming this structure, the number of sample years associated with frequencies of 1, 2, 3, 4, and 5 years were 21, 11, 8, 6, and 5, respectively.

Programs 10 years in length spanned the same range of number of surveys. Number of sites varied between 10 and 100 in increments of 10. Designs with sampling every one, two, three, or five years were simulated. Sampling every three years again implied incomplete coverage of the program length, with the final sample year occurring in year nine. This problem was avoided by sampling in the 10^{th} instead of the ninth year. A sampling frequency of four years was not simulated because it did not cover the entire program length and achieved the same number of sample years as programs with sampling frequency equal to five years. Given these adjustments, the number of sample years associated with frequencies of one, two, three, and five years were 11, six, four, and three, respectively. All 10 year programs assumed 16 point count stations. This reflected preliminary results from 20 year program power analyses that indicated point count stations should be maximized.

For population monitoring, the ability of these sampling designs to detect –3% trends was then tested. 3% per year was selected because it is the effect size goal of the Alberta Forest Biodiversity Monitoring Program, a large-scale biodiversity monitoring program currently being developed for Alberta's forested regions (Schneider et al. 1999). Over 20 years, this effect size equates to a 45% decrease. Over 10 years, the total decline is 26%. When simulating community metric trajectories, 1% per year was selected as the effect size. This is equivalent to total declines of 10 and 18% for ten and twenty year programs, respectively. A reduced effect size was used because a given reduction in a community metric, such as species richness, implies greater degradation of the community than the same reduction in a population.

Sensitivity Analysis

To evaluate the influence of each species parameter (α_0, τ_1^2, τ_2^2, τ_3^2) on power, the following sensitivity analysis was performed. For each parameter, power to detect -3% trends over 20 years (alpha = 0.2) was estimated at maximum and minimum levels of the parameter as estimated across species and effort levels. The other three species parameters were kept constant at their average levels across species and effort levels.

Number of sites and sampling frequency were kept constant at minimum levels (10 sites sampled every five years) to increase sensitivity of power to species parameters.

Modeling the Relationship Between Power and Sample Effort

Comprehensive graphical presentation of the influence of sample effort on power was difficult due to the numerous scenarios simulated and sample effort parameters involved. Instead, two models (one for each program length) were developed for each species and community metric that expressed power as a function of number of sites, number of sample years, number of point count stations, and number of surveys. This was accomplished using multiple regression, with power as the dependent variable and sample effort parameters as covariates.

Residual analysis indicated a linear model was unsuitable, due to the asymptotic nature of power as it approaches the 100% maximum. Instead, the following nonlinear model was used:

$$y = 1 - \exp(-1 * (a_0 + a_1 s + a_2 f + a_3 p + a_4 r + a_5 sf + a_6 sp + a_7 sr + a_8 fr + a_9 fp + a_{10} pr)) \quad \textbf{(3.1)}$$

where

y = power

$a_{0\text{-}10}$ = model coefficients

s = number of sites

f = number of sample years

p = number of point count stations

r = number of surveys.

The model was fit to the power estimates by minimizing the residual sum of squares using procedure nls in S-Plus (MathSoft 1999). Residual analysis suggested no violations of normality and homogeneity assumptions. For 10 year programs, terms involving number of point count stations were eliminated from the model because this effort parameter was fixed at 16, producing the following model:

$$y = 1 - \exp(-1 * (a_0 + a_1 s + a_2 f + a_3 r + a_4 sf + a_5 sr + a_6 fr)) \quad \textbf{(3.2)}$$

Estimating Rates of Power Gain with Effort

To estimate the rate of power gain with respect to each sample effort parameter, partial derivatives of Equations 3.1 and 3.2 were calculated (Appendix 2). For each species/community metric and program length combination, five partial derivative solutions were obtained, one for each of five levels of power: <= 60%; > 60% and <= 70%; >70% and <= 80%; >80% and <= 90%; and >90%. This was accomplished by solving the partial derivative using mean levels of each sample effort parameter across sampling designs that achieved power within the interval of interest, and coefficient values as estimated using nonlinear regression. This produced estimates of the average rate at which power increased with respect to each effort parameter, within the five intervals, for each species and community metric. To evaluate whether partial derivative solutions were significantly different than zero, z statistics were calculated using standard errors estimated following the Krinsky-Robb method (Krinsky and Robb 1986). This method uses simulation to estimate the standard error of a nonlinear function of parameter estimates from a model. Here, partial derivatives were nonlinear functions of the estimated coefficient vectors (a_0-a_{10}) of Equations 3.1 and 3.2. To estimate the standard error of a partial derivative, 10,000 coefficient vectors were generated by drawing random vectors from a multivariate normal distribution with mean and variance equal to the coefficient vector and variance-covariance matrix as estimated using nonlinear regression. 10,000 partial derivative estimates were then calculated using the 10,000 coefficient vectors. These estimates were used to approximate the distribution of the partial derivative, and standard error of the partial derivative was estimated as the square route of the variance of the 10,000 partial derivative estimates. The method requires the assumption that parameters are distributed multivariate normal with mean and variance-covariance equal to the mean and variance-covariance estimated using nonlinear regression. Normal distribution of the residuals, as was the case here, validates this assumption (Myers 1990).

Results

Power Estimates

Power estimates refer to power to detect –3% per year trends in species abundance and –1% per year trends in community metrics. In all cases, a power goal of 90% and a significance goal of 80% are assumed.

20 Year Monitoring Programs

Because all simulated designs achieved greater than 90% power for White-throated Sparrow, species richness, and the Shannon-Weiner diversity index, these monitoring targets are not included in results that follow. Figure 3.1 illustrates the sensitivity of power to sample effort parameters across species. Each figure varies a single effort parameter across its range, while keeping the other effort parameters constant at minimum levels. Minimum levels were used in an attempt to isolate the influence of the effort parameter of interest. In general, power increased with each sample effort parameter. Power levels differed among species, which could be partitioned into three groups. Species with highest power were the Black-throated Green Warbler and Yellow Warbler. Species with lowest power were the Pileated Woodpecker, Gray Jay, and White-breasted Nuthatch. The Brown-headed Cowbird and Black-capped Chickadee exhibited slightly higher power levels than these species. Power to detect trends in species richness of the ground-nesting guild was higher than for all species, except the White-throated Sparrow.

10 Year Monitoring Programs

The power of sampling designs to detect trends over 10 years was lower than the power of the same designs over 20 years (e.g. Figure 3.2). As with 20 year programs, power estimates differed across species (Figure 3.3). None of the simulated sampling designs achieved 90% power for Pileated Woodpecker, Gray Jay, and White-breasted

Nuthatch. In contrast, all simulated designs detected trends with greater than 90% power for the Shannon-Weiner diversity index, which was not included in Figure 3.3.

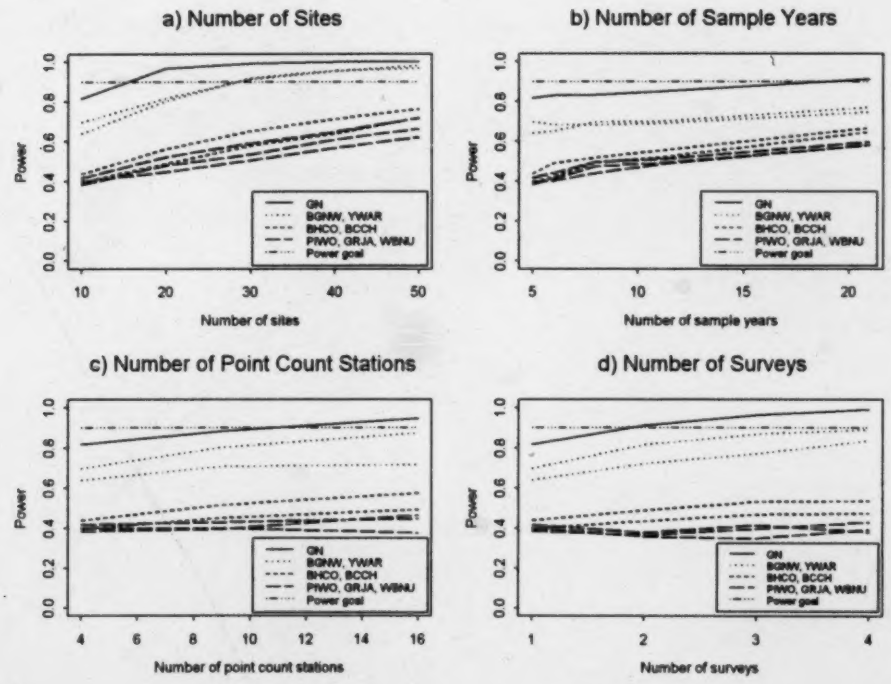

Figure 3.1. Response of power to a) number of sites, b) number of sample years, c) number of point count stations, and d) number of surveys. Unless otherwise stated in the figure, sample effort is fixed minimum levels. The four letter codes identify seven species and one community metric (BCCH = Black-capped Chickadee, BGNW = Black-throated Green Warbler, BHCO = Brown-headed Cowbird, GRJA = Gray Jay, PIWO = Pileated Woodpecker, White-breasted Nuthatch, YWAR = Yellow Warbler, GN = species richness of the ground-nesting guild). For each collection of species in the legends, species are listed from highest to lowest power. For example, for the low-power species group, PIWO has the highest power, followed by GRJA and WBNU.

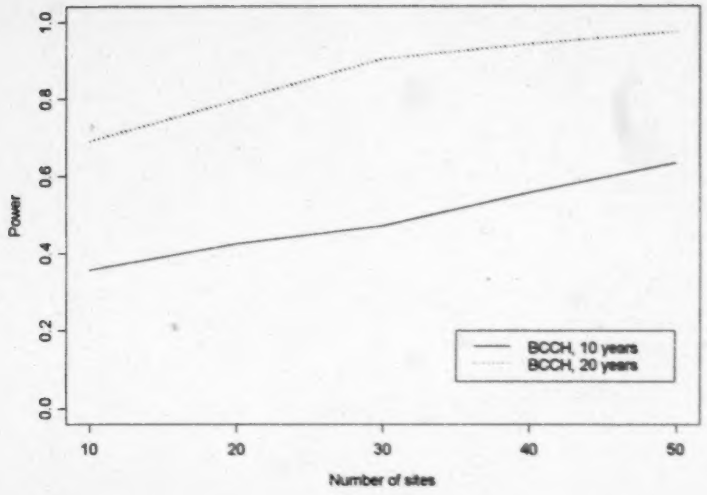

Figure 3.2. Responses to number of sites of power to detect a -3% per year trend in Black-capped Chickadee (BCCH) abundance over 10 and 20 years. The response surfaces refer to effort levels of sample frequency = 5, point count stations = 16, surveys = 1.

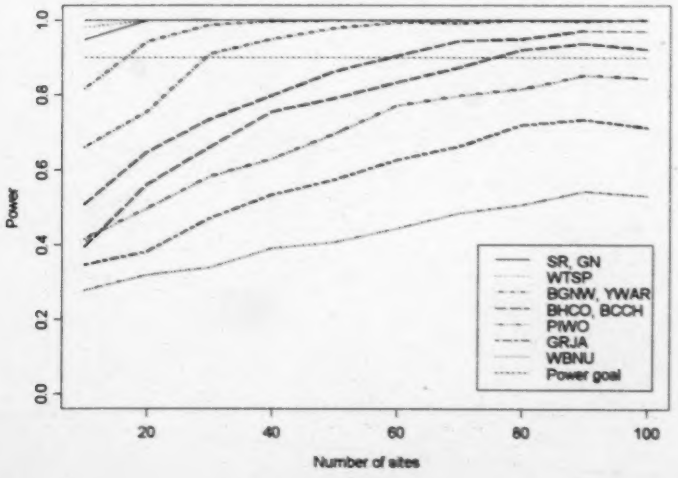

Figure 3.3. Response to number of sites of power to detect -3% per year trends in species' abundance and -1% per year trends in community metrics over 10 years. The response surfaces refer to effort levels of sample frequency = 1, point count stations = 16, surveys = 4. Target codes as in Figure 3.1, and WTSP = While-throated Sparrow, SR = species richness.

Sensitivity Analysis

Power was highly sensitive to within-site temporal variance, and moderately sensitive to between-site variance in initial abundance (Table 3.3). Between-site variance in trend also influenced power, although to a still lesser extent. Power was not sensitive to abundance.

Parameter	α_0	τ_1^2	τ_2^2	τ_3^2
Minimum parameter estimate	0.049	0.276	0.557	0.139
Power at minimum parameter estimate	0.558	0.998	0.681	0.613
Maximum parameter estimate	20.094	2.449	2.254	0.487
Power at maximum parameter estimate	0.548	0.447	0.565	0.558
Power at minimum minus power at maximum	0.010	0.551	0.116	0.055

Table 3.3. Results of a sensitivity analysis evaluating the influence of species' parameters (α_0, τ_1^2, τ_2^2, and τ_3^2) on power to detect trend (see text for details).

Power Models

Power estimates were summarized by modeling the effect of sample effort on power, using Equation 3.1. The resulting models explained much of the variability in power estimates (Table 3.4). Results were not modeled for species or community metrics for which all sampling designs achieved greater than 90% power or for which no sampling designs achieved greater than 90% power. This excluded species richness, the Shannon-Weiner diversity index, and the White-throated Sparrow for 20 year programs, and the Shannon-Weiner diversity index, Pileated Woodpecker, Gray Jay, and White-breasted Nuthatch for 10 year programs.

	BCCH	BGNW	BHCO	GRJA	PIWO	WBNU	WTSP	YWAR	GN	SR
10 years	0.965	0.990	0.963	NA	NA	NA	0.992	0.970	0.979	0.800
20 years	0.974	0.984	0.976	0.940	0.950	0.959	NA	0.977	0.960	NA

Table 3.4. Coefficients of determination (R^2) for nonlinear regressions relating sample effort parameters to power. Target codes as in Figure 3.1, and WTSP = White-throated Sparrow, SR = species richness.

Rates of Power Gain with Effort

Partial derivatives for power decreased over the five power intervals (Figure 3.4), indicating the rate of increase in power due to increased sample effort was smaller at higher power levels. Average rates of power gain for designs achieving between 80 and 90% power, and their p-values, are provided in Table 3.5 (20 year programs) and Table 3.6 (10 year programs). This power interval is emphasized to evaluate the influence of sample effort parameters on power within the proximity of the 90% power goal. The rates of power gain represent the increase in power achieved by increasing the number of sites, sample years, point count stations, or surveys by one, within the range of effort levels simulated (Table 3.2). Rates of power gain with point count stations were not calculated for 10 year programs because only one level of this effort parameter (16) was simulated.

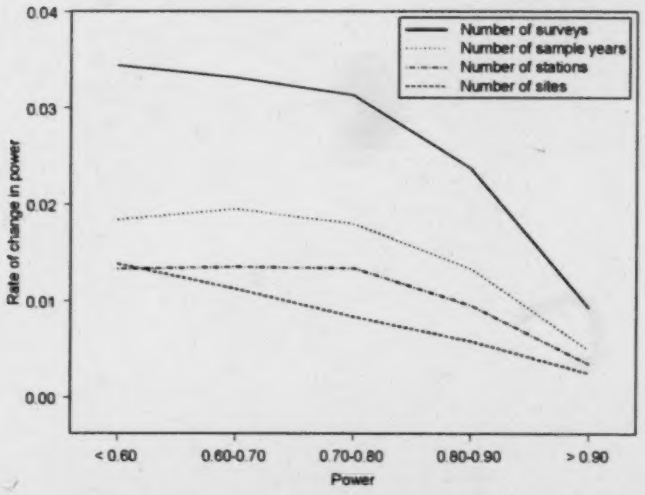

Figure 3.4. Rate of change in power to detect -3% trends in BCCH over 20 years with respect to four sample effort parameters, averaged across designs achieving one of five levels of power.

Species	Number of sites	Number of sample years	Number of point count stations	Number of surveys
BCCH	0.0058 (0.000)	0.0133 (0.000)	0.0095 (0.000)	0.0237 (0.000)
BGNW	0.0114 (0.000)	0.0016 (0.000)	0.0169 (0.000)	0.0643 (0.000)
BHCO	0.0062 (0.000)	0.0100 (0.000)	0.0104 (0.000)	0.0296 (0.000)
GRJA	0.0054 (0.000)	0.0145 (0.000)	0.0046 (0.000)	-0.0003 (0.891)
PIWO	0.0052 (0.000)	0.0142 (0.000)	0.0068 (0.000)	0.0224 (0.000)
WBNU	0.0058 (0.000)	0.0137 (0.000)	0.0016 (0.000)	0.0026 (0.147)
YWAR	0.0156 (0.000)	0.0043 (0.000)	0.0053 (0.000)	0.0470 (0.000)
GN	0.0265 (0.000)	0.0075 (0.000)	0.0125 (0.000)	0.1431 (0.000)
Average	0.0103	0.0098	0.0084	0.0416

Table 3.5. Rates of change (and p-values) in power to detect -3% per year trends in species' abundance and -1% per year trends in community metrics over 20 years, with respect to four sample effort parameters. Species and community metric codes as in Figure 3.1.

Species	Number of sites	Number of sample years	Number of surveys
BCCH	0.0030 (0.000)	0.0242 (0.000)	0.0386 (0.000)
BGNW	0.0076 (0.000)	0.0010 (0.027)	0.0844 (0.000)
BHCO	0.0031 (0.000)	0.0240 (0.000)	0.0525 (0.000)
YWAR	0.0041 (0.000)	0.0267 (0.000)	0.0618 (0.000)
WTSP	0.0216 (0.000)	0.0034 (0.000)	0.0742 (0.000)
GN	0.0109 (0.000)	0.0098 (0.000)	0.0655 (0.000)
SR	0.0245 (0.000)	0.0140 (0.000)	0.0407 (0.000)
Average	0.0107	0.0172	0.0597

Table 3.6. Rates of change (and p-values) in power to detect -3% per year trends in species' abundance and -1% per year trends in community metrics over 10 years, with respect to three sample effort parameters. Species and community metric codes as in Figure 3.1, and WTSP = White-throated Sparrow, SR = species richness.

Discussion

Power to Detect Population Trends

Because power to detect trend differed across species, sampling designs facilitating adequate trend detection (90% power) for some species achieved much lower power for others. For example, designs associated with between 88 and 92% power for detection of –3% per year trends in Yellow Warbler abundance over 20 years detected trends in White-breasted Nuthatch abundance with, on average, 59.3% power. Because sampling

design requirements differ greatly across species, monitoring programs should be designed to achieve adequate trend detection in the species or metric of interest that is associated with the lowest power. This will ensure adequate data are collected for the full range of species or metrics targeted by the program.

Evaluating how species' characteristics were related with power estimates provides insight into which types of species will require high sample effort. Power to detect trend was greatest for species exhibiting low variance. This negative relationship was strongest for within-site temporal variance, lesser for between-site variance in initial abundance, and negligible for between-site variance in trend. The minimal influence of between-site variance in trend on power is not surprising, given the small range of this parameter across species. Although the sensitivity analysis indicated abundance did not influence power, power to detect trend was lower for rarer species, such as the White-breasted Nuthatch. This tendency, which is explained by the negative relationship between abundance and variance parameters (Chapter 2), implies population monitoring programs will require high sample effort if trends in rare species are to be detected.

Power to Detect Community Trends

The high variability of populations has lead some to promote monitoring of community metrics in favor of individual species' populations (e.g. Karr and Chu 1997). Results presented here support this stance, as power to detect trends was dramatically higher for community metrics than for most populations. The comparative ease with which trends were detected in community metrics suggests community monitoring may be an efficient way to track ecosystem status. However, community metrics were associated with high power because of their low variability. This same low variability makes the metrics insensitive to disturbance as compared to populations. For example, a much larger disturbance is required to reduce species richness by 10% than a population by the same amount. To account for this insensitivity, effect size for community metrics was set at –1% per year as compared to –3% per year for populations. This reduction may not account for the decreased sensitivity of community metrics, and further research into the sensitivity of community metrics to disturbance is required. The high power

associated with richness of the ground-nesting guild and the Shannon-Weiner diversity index is encouraging, as these metrics provide information likely to be of more use than total number of species alone. Community composition, which can be monitored using guilds, is likely to be more sensitive to disturbance than total number of species (Croonquist and Brooks 1991), as is distribution of individuals across species (Magurran 1988), which can be monitored using diversity indices.

The Effect of Program Length on Power

As has been observed in other power analyses (e.g. Lewis and Gould 2000), program length dramatically affected power. Low power achieved by 10 year programs as compared with 20 year programs is not surprising, given the reduction in cumulative effect size. For example, a –3% decline over 10 years accounts for an overall decline of 26%, as compared to an overall decline of 45% over 20 years. If monitoring programs are to act as early warning signals of the degradation of ecosystems, trends must be detected over relatively short time intervals (i.e. 10 years or less). Results presented here indicate this will require high levels of sample effort, especially for rare species.

The Effect of Sample Effort on Power

Power increased with sample effort at a decreasing rate at higher power levels, reflecting the asymptotic relationship between sample effort and power. This implies effort required to increase power from moderate levels, such as 70%, to relatively high levels, such as 90%, will be substantial. For sampling designs achieving between 80 and 90% power, the rate of increase in power caused by increased effort was significantly greater than zero ($p < 0.05$) for all species, community metrics, program lengths and effort parameters, with the exception of rate of increase in power due to surveys for the Gray Jay and White-breasted Nuthatch for 20 year programs. The rate of power gain associated with point count stations was high, considering the minimal amount of effort required to sample one additional point count station as compared to the effort required to sample one additional site, survey, or year. High point count station density would

therefore appear to be a component of an efficient sampling design. Based on this result, number of point count stations was fixed at 16 when simulating 10 year monitoring programs. Average rates of increase in power with the three remaining sample effort parameters were greater for 10 year programs than 20 year programs. The general pattern across effort parameters remained, however, with the rate of increase due to surveys being greater than those due to number of sites and sample frequency. Despite this pattern, it is not clear how effort should be divided between these three parameters in the absence of cost information, because rates of increase in cost will also differ across effort parameters.

Rates of power increase differed across species. In general, the rate at which power increased with sample frequency was greater for temporally variable species. With respect to 20 year programs, the correlation between the ranking of species by within-site temporal variance and the ranking of species by rate of increase in power with frequency was 0.893 ($p = 0.007$). The rate at which power increased with number of sites also appeared to be related to within-site temporal variance, although in a negative fashion. Again with respect to 20 year programs, the correlation between rankings by within-site temporal variance and rate of increase in power with sites was –0.808 ($p = 0.028$). These relationships suggest optimal allocation of sample effort may differ depending on species' within-site temporal variance in the following manner. When monitoring species with high within-site temporal variance, typically those with low abundance, effort should be allocated to sampling frequency in favor of number of sites. When monitoring species with low within-site temporal variance, on the other hand, more emphasis should be placed on sampling numerous sites.

Within-site sample effort tended to decrease sample error (Chapter 2), which in turn increases power to detect trends (Steidl et al. 1997). It was not surprising, therefore, that the rate at which power to detect trend in a species increased with within-site effort depended on the degree to which sample error was influenced by within-site effort. For 20 year programs, the correlation between rankings of species by the effect of number of surveys on sample error (Chapter 2) and by the rate of increase in power with surveys was 0.929 ($p = 0.003$). In addition, the only two rates of increase in power that were not significant (increase in power due to surveys for Gray Jay and White-breasted Nuthatch)

can be explained by the negligible effect of surveys on sample error. Similarly, the correlation between rankings of species by the effect of number of stations on sample error (Chapter 2) and by the rate of increase in power with stations was 0.893 ($p = 0.007$).

Although the differences in rates of power increase suggest strategies to increase sampling efficiency, some minimum level of each effort parameter will be required to detect trends with high power, precluding maximization of one parameter in favor of the other. For example, although rate of increase in power due to number of sample years is greater than rate of increase due to number of sites for the White-breasted Nuthatch, no sampling design with less than 40 sites detected trends in this species with at least 90% power over 20 years. In addition, it is difficult to determine how effort should be allocated in the absence of cost information. For example, one would expect the cost of increasing the number of sample years by one to be greater than the cost of increasing the number of sites by one, which suggests allocation of effort to number of sites in favor of number of sample years may be advantageous even when the rate of increase due to number of sample years is greater.

Implications of Simulation Assumptions

This study evaluated power to detect –3% exponential trends in species abundance. Smaller effect sizes will require increased sample effort to achieve equivalent power (Gibbs and Melvin 1997, Gibbs et al. 1998, Lewis and Gould 2000). Negative trends were simulated based on the assumption that their detection will be of greater importance than positive trends. Positive trends will also be of interest, especially for species that respond positively to forest disturbance (e.g. Brown-headed Cowbird). However, due to their assumed exponential shape, positive trends will be associated with larger overall changes and will be easier to detect (Lewis and Gould 2000). Therefore, power to detect positive trends will be greater than estimates presented here. Exponential trends were simulated based on the assumption that populations experience constant rates of decline (Caughley and Sinclair 1994). If trends are linear, overall declines will be larger than if trends are exponential, and associated power will be higher. Simulation of negative exponential trends, therefore, provides conservative estimates of power as compared to

simulation of negative linear trends. If, however, trends follow an alternative, irregular trajectory, estimates presented here will overestimate power to detect trend (Gibbs et al. 1998). Irregular trends would also make the assumed practice of sampling all sites in common years inappropriate. An alternating design, where at least some sites are monitored in any given year, would have increased ability to detect irregular patterns such as dramatic, but short-term, trends.

Significance was set at 80% when estimating trend. 80% significance implies a 20% type I error rate, where type I error is the probability the null hypothesis (no trend) is accepted when the null is true. A check of type I error indicated moderate exaggeration. Type I error rate was estimated by simulating 1000 0% trends and determining the proportion of trends for which the null was rejected. Average type I error (and 95% confidence interval) across 20 year sampling designs was 26.2% (25.8, 26.6) for a ubiquitous species (Yellow Warbler) and 26.4% (25.9, 26.9) for a variable species (Pileated Woodpecker). For 10 year designs, average type I error increased slightly to 27.7% (26.8, 28.6) and 27.2% (26.1, 28.2) for the Yellow Warbler and Pileated Woodpecker, respectively. Excessive type I error was likely due to two tendencies of route regression: to underestimate the standard deviation of trend estimates (Thomas 1997) and to estimate trend with positive bias (Geissler and Link 1988). The effect on power of type I error exceeding nominal levels is difficult to interpret because the implications of the two tendencies differ. Underestimated standard deviation in trend would cause optimistic power estimates (Thomas 1997). On the other hand, in the presence of negative trend, positive bias in trend estimates would cause power to be underestimated. Simulations described above suggested positive bias was present, although small. For the Pileated Woodpecker, average magnitude of trend detected in the absence of real trend was 0.253% ($p = 0.000$) for 20 year designs and 0.676% ($p = 0.000$) for 10 year designs. As a result, no attempt was made to correct for exaggerated type I error by altering the significance level when estimating power.

Summary

Large-scale avian sampling design was evaluated using a simulation approach that estimated power of a range of sampling designs to detect trends in eight bird species and three bird community metrics. Power varied across species, with variable species being associated with lower power. Because variable species, which also tend to be rare, are likely to be of monitoring interest, population monitoring programs may require high levels of effort to achieve adequate power. Detection of trend in community metrics, on the other hand, was much more powerful. Comparison of power to detect trend over 10 and 20 year periods indicated detection of trend over relatively shorter time periods will require dramatically more sample effort. To evaluate the effect of individual sample effort parameters on power, the relationship between power estimates and sample effort parameters was modeled using nonlinear multiple regression. Calculation of the models' partial derivatives facilitated estimation of the rates at which power increased with increases in each effort parameter. Power increased rapidly with point count stations, relative to the low effort required to sample more point count stations. Efficiency of allocation of effort to other effort parameters is unclear in the absence of cost information. Integration of rates of power increase with rates of cost increase would facilitate identification of cost-effective strategies for the allocation of effort to sample effort components.

Chapter 4

Cost-effective Sampling Strategies for Detecting Long-term Avian Trends

Introduction

In recognition of extensive monitoring requirements coupled with limited financial resources (Mac Nally 1997), the need to monitor ecosystems cost-effectively has been established (Usher 1989, Committee on Environment and Natural Resources 1996, Schneider et al. 1999). Indeed, program cost is one of the most important factors considered during monitoring program development (Samways et al. 1995), and realistic cost levels are essential for long-term program viability. Funding limitations impose tradeoffs between the allocations of effort to sample effort components (Committee on the Environment and Natural Resources 1997), which compete for monitoring resources (Millard and Lettenmaier 1986). This implies cost-effective monitoring design requires an understanding of the efficiency of allocating effort to sample effort components.

Evaluation of effort efficiencies necessitates estimation of the influence of sample effort on both cost and monitoring program effectiveness. The relative benefit of allocating effort to each sample effort component may not be equal (Bernstein and Zalinski 1983, Millard and Lettenmaier 1986). Therefore, maximizing data quality under cost constraints requires knowledge regarding the efficiency of each effort component, where efficiency refers to the rate at which data quality (power) increases as compared to the rate at which cost increases as effort is allocated to an effort component.

Previous investigations of cost-effective monitoring design addressed how effort should be allocated for relatively short-term impact assessment (e.g. Bernstein and Zalinski 1983, Millard and Lettenmaier 1986, Mac Nally 1997). Here, cost-effective design with respect to avian monitoring programs is evaluated, where the goal is detection of long-term trend. The sampling design for a large-scale avian monitoring program was described in Chapter 2. Assuming this design framework, the task of

designing a monitoring program can be simplified to determining how effort should be allocated to sample frequency, site density, and effort per site, such that adequate power is achieved in an efficient manner. This study focuses on the bird community due to its suitability for monitoring (Steele et al. 1984, Furness et al. 1993, Moors 1993), prevalence as a monitoring target (e.g. Dunn et al. 1993), and the availability of data required for the analysis.

In this chapter, a relationship between program cost and sample effort is established, and the rate at which cost changes with respect to each effort parameter is estimated. These results are related to avian monitoring program power estimates, facilitating a comparison of the relative efficiency of allocating effort to each sample effort parameter.

Methods

Evaluation of the efficiency of sample effort allocation involved two steps. First, costs of alternative sampling designs described in Chapter 3 were calculated according to a cost model. Second, estimates of sampling design cost were integrated with estimates of sampling design power. This involved determining the least expensive sampling design achieving 90% power for each species and community metric of interest, and comparing rates of increase in cost to rates of increase in power across sample effort components. Multiple monitoring targets were included in the analysis to facilitate a comparison of sampling efficiencies across targets. Populations of several species were considered to represent a range of variance characteristics (Chapter 2). The variance categories (and the relevant species) were ubiquitous (White-throated Sparrow and Yellow Warbler), spatially variable (Black-throated Green Warbler and Brown-headed Cowbird), temporally fluctuating (Gray Jay and Black-capped Chickadee), and irruptive (White-breasted Nuthatch and Pileated Woodpecker). Three types of community metrics were also evaluated: total species richness, species richness of a guild (ground-nesting species richness), and a diversity index (Shannon-Weiner index).

Estimating Sampling Design Cost

To estimate the cost of sampling designs, a cost model was developed that divided cost into two components: field costs and aerial access costs. Total cost was the sum of the two components. Annual cost was obtained by dividing total cost by the maximum possible number of sample years (21 for 20 year programs and 11 for 10 year programs). Except where noted otherwise, estimates of parameters used in cost calculations are those assumed by the Alberta Forest Biodiversity Monitoring Program, a large-scale biodiversity monitoring program currently in the design phase (D. Farr, pers. comm.).

Field Costs

Field costs include labor, equipment, and terrestrial access to sites. Sampling of sites is conducted by crews, each comprising of two individuals. Each crew samples 30 days during the two month sampling period (May and June), accounting for days off and bad weather. Because each crew can sample and enter data from one site per day, regardless of whether number of point count stations is four, nine, or sixteen (F. Schmiegelow, pers. comm.), each crew samples 30 sites per year. Sampling more than one site per day is not possible due to the restricted time interval during which sampling can occur (the morning) and the time required to access spatially segregated sites. Therefore, the number of crews required equals the number of sites requiring sampling (number of sites multiplied by number of surveys) divided by 30. Cost of sampling then equals the number of crews multiplied by the cost per team over the two month period ($26,667). In addition to sampling visits, sites are visited once in the year prior to commencement of sampling to locate the site and identify point count stations. Cost per site for this activity is $1,333, excluding aerial access cost. According to this scenario:

$$\text{Field costs} = \text{years} * (\text{visits}/\text{crew}_{\text{sites}}) * \text{crew}_{\text{cost}} + \text{sites} * \text{initial}_{\text{cost}} \quad \textbf{(4.1)}$$

where,

visits=number of site visits per sample year (number of sites * number of surveys)

years = number of sample years

sites = number of sites

$crew_{sites}$ = sites sampled per crew per year (30)

$crew_{cost}$ = cost per crew per year, including labor, equipment, food, vehicles, and gas ($26,667)

$initial_{cost}$ = cost to visit each site prior to commencement of sampling ($1,333).

Aerial Access Cost

Although field costs includes truck, quad, and gas costs, and therefore accounts for terrestrial access to sites, it does not account for aerial access costs. Over the duration of the monitoring program, aerial access cost is calculated according to Equation 4.2.

$$\text{Aerial access cost} = (\text{sites} + \text{years} * \text{visits}) * (heli_{prop} * heli_{time} * heli_{cost}) \quad \textbf{(4.2)}$$

where,

$heli_{prop}$ = proportion of sites requiring aerial access

$heli_{time}$ = helicopter time required per site visit

$heli_{cost}$ = hourly helicopter cost ($750)

Equation 4.2 assumes the proportion of sites requiring aerial access is independent of the year of the monitoring program. However, the proportion will decrease as road density in northern Alberta increases with increasing development. This relationship is modeled by assuming the proportion will decrease by a constant amount each year, as described in Equation 4.3:

$$heli_{prop}(t) = heli_{prop}(0) - prop_{decrease} * t \quad \textbf{(4.3)}$$

where,

$heli_{prop}(t)$ = proportion of sites that require aerial access in year t

$heli_{prop}(0)$ = proportion of sites requiring aerial access in year 0 of the monitoring program (50%)

$prop_{decrease}$ = decrease in the proportion of sites requiring aerial access with each additional year.

Because Equation 4.3 implies cost must be calculated separately for each year of the monitoring program (due to the variable t), it could not be used when estimating rates of total cost increase using partial derivatives. To avoid this problem, aerial access cost was calculated using Equation 4.2, with $heli_{prop}$ set at the average $heli_{prop}(t)$ over the entire

monitoring period. For example, in reference to 20 year monitoring programs, if $prop_{decrease}$ equaled 1%, $heli_{prop}$ was set at 40%. The rate at which the proportion of sites requiring aerial access will decrease was unknown, requiring cost to be calculated under three scenarios: no decrease ($prop_{decrease} = 0$); moderate decrease ($prop_{decrease} = 1.25\%$); and large decrease ($prop_{decrease} = 2.5\%$). Over 20 years, these decreases accumulated to 0, 25%, and 50%, respectively.

Equation 4.2 also assumes helicopter time required per site visit is independent of the number of site visits required per sample year. A more likely scenario is the required helicopter time per visit will decrease as site visits per sample year increase, due to increased efficiency. Again, a constant decrease was assumed, according to Equation 4.4.

$$heli_{time}(v) = heli_{time}(10) - time_{decrease} * (sites - visits_{min}) \quad \textbf{(4.4)}$$

where,

$heli_{time}(10)$ = helicopter time required per site visit when number of site visits is at its minimum (2 hours)

$time_{decrease}$ = decrease in helicopter time required per site visit with each additional required site visit

$visits_{min}$ = minimum number of site visits per sample year (10).

Because the rate of decrease was again unknown, cost was calculated under three scenarios: no decrease ($time_{decrease} = 0$); medium decrease ($time_{decrease} = 45$ seconds); and large decrease ($time_{decrease} = 90$ seconds). According to these scenarios, if number of sites were increased from 10 to 50, the decrease in helicopter time required per site would be 0, 30, and 60 minutes.

Estimating Least Expensive Effective Designs

Power to detect –3% per year exponential trends in species and –1% per year exponential trends in each community metric was estimated for each sampling design scenario using Monte Carlo simulations (Chapter 3). These estimates were then combined with cost estimates to determine the least expensive sampling design detecting trends with at least 90% power at 80% significance for each metric, for programs 10 and

20 years in length. To estimate the robustness of the least expensive effective design (LEED), designs within $5,000 of its annual cost were investigated.

Estimating Relative Efficiency of Effort Parameters

Determining the relative efficiency of effort parameters required estimating the rate at which power and cost increased with respect to each sample effort component. Rates of power gain for each metric were estimated by taking partial derivatives of models expressing the relationship between sample effort and power, and solving using average levels of effort parameters across designs achieving between 80 and 90% power (Chapter 3). Rates of cost increase were estimated by taking partial derivatives of the cost model (Appendix 3), and solving using the same effort parameter levels used when solving power partial derivatives. This provided estimates of the average rates at which power and cost increased with each effort parameter, for sampling designs achieving between 80 and 90% power.

The cost model assumed increases in point count stations incurred no additional cost. This assumption, combined with positive power partial derivatives for all metrics (Chapter 3), implied rates of increase in power were always greater than rates of increase in cost for point count stations. To compare the power and cost rates of increase associated with the remaining effort parameters, their partial derivatives were converted into proportions by dividing power or cost partial derivatives related to each effort parameter by the sum of power or cost partial derivatives across the effort parameters. This was required because cost partial derivatives were much greater than power partial derivatives. For a given effort parameter, if the power proportion was greater than the cost proportion, the rate at which power increased by allocating effort to that parameter was greater than the rate at which cost increased, relative to the other parameters. This indicated efficiency. If, on the other hand, the cost proportion was greater, the cost of allocating effort to that parameter was greater than the power gains achieved, as compared to the other parameters. This indicated inefficiency.

Because power partial derivatives were derived from a regression model which only approximately represented the relationship between power and sample effort parameters,

uncertainty was associated with each power proportion. To account for this uncertainty, 95% confidence intervals for the power proportions were calculated using the Krinsky-Robb method (see Chapter 3).

Results

20 Year Monitoring Programs

Cost Estimates

Program cost was estimated using the three levels of decline in the proportion of sites requiring helicopter access and amount of helicopter time required per site visit described in the Methods. Average annual costs across programs for the three levels are presented in Table 4.1. Not surprisingly, program cost decreased as the rate of decline in helicopter use increased, as did the proportion of cost allocated to aerial access of sites.

Scenario	No decrease	Moderate decrease	Dramatic decrease
$time_{decrease}$	0	45 seconds	90 seconds
$prop_{decrease}$	0	1.25%	2.5%
Mean annual cost	$49,017	$40,870	$35,000
Mean % of cost due to aerial access	30.3%	17.4%	3.4%

Table 4.1. Mean 20 year program annual cost and mean percentage of cost accountable to aerial access, for each aerial access cost decline scenario.

Least Expensive Effective Designs

As expected, power to detect trends increased as the cost of monitoring programs increased (e.g. Figure 4.1). The relationship was highly nonlinear, however, which resulted in a large range of costs being associated with sampling designs achieving 90% power or greater. The least expensive sampling design achieving 90% power for a given species or community metric was not affected by the rate at which helicopter access cost decreased. The cost of the LEED decreased, however, with higher rates of decline in

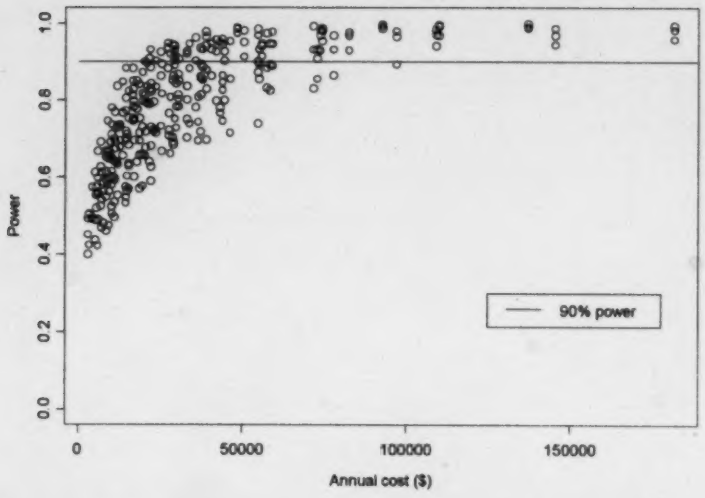

Figure 4.1. Power to detect –3% per year exponential trends in Black-capped Chickadee abundance over 20 years versus annual program cost.

helicopter cost. The annual costs and sample effort levels for LEED's are given in Table 4.2. In general, species exhibiting higher within-site temporal variance were associated with higher minimum costs (Figure 4.2). For each sample effort parameter, the minimum and maximum levels observed across designs achieving at least 90% power and costing within $5,000 of the annual cost of the LEED are also presented. A large range of sample effort levels were observed within the cost margin, indicating LEED's were not robust.

Relative Efficiency of Effort Parameters

The cost partial derivatives for each effort parameter, solved using average levels of effort parameters achieving between 80 and 90%, are given in Appendix 3. Power partial derivatives are presented in Chapter 3. Cost proportions were robust to aerial access cost decline scenarios. The average difference (and standard deviation) in cost proportions between the maximum and minimum scenarios were 0.0002 (0.0003), 0.0012 (0.0019), and –0.0015 (0.0022) for sampling frequency, number of surveys, and number of sites,

Species	Cost (scenario 1)	Cost (scenario 2)	Cost (scenario 3)	Number of sites	Sampling frequency	Number stations	Number of surveys
BCCH	$28,042	$23,533	$20,363	40 (30,40)	5 (5,2)	16 (16,16)	2 (1,3)
BHCO	$18,412	$15,600	$13,591	40 (40,50)	4 (5,4)	16 (16,16)	1 (1,1)
BGNW	$7,010	$6,118	$5,225	10 (10,20)	5 (5,3)	9,16 (9,16)	2 (1,3)
GRJA	$38,061	$31,477	$27,347	50 (50,50)	2 (2,2)	9,16 (9,16)	1 (1,1))
PIWO	$50,098	$41,281	$35,813	50 (40,50)	5 (5,1)	16 (4,16)	3 (1,4)
WBNU	$54,523	$45,349	$38,987	40 (40,50)	1 (2,1)	4,9,16 (4,16)	1 (1,1)
WTSP	$4,001	$3,555	$3,108	10 (10,20)	5 (5,2)	4,9,16 (4,16)	1 (1,2)
YWAR	$9,206	$8,023	$6,974	20 (10,30)	4 (5,4)	16 (4,16)	1 (1,4)
GN	$4,001	$3,555	$3,108	10 (10,20)	5 (5,2)	16 (4,16)	1 (1,2)
SR	$4,001	$3,555	$3,108	10 (10,20)	5 (5,2)	4,9,16 (4,16)	1 (1,2)
SW	$4,001	$3,555	$3,108	10 (10,20)	5 (5,2)	4,9,16 (4,16)	1 (1,2)

Table 4.2. Least expensive effective designs for 20 year programs. The four letter target codes in the legends identify eight species and three community metrics (BCCH = Black-capped Chickadee, BGNW = Black-throated Green Warbler, BHCO = Brown-headed Cowbird, GRJA = Gray Jay, PIWO = Pileated Woodpecker, WBNU = White-breasted Nuthatch, WTSP = White-throated Sparrow, YWAR = Yellow Warbler, GN = species richness of the ground-nesting guild, SR = total species richness, SW = Shannon-Weiner diversity index). Costs are presented for each aerial access cost decline scenario. In brackets next to the effort levels for each minimum cost effective design are the minimum and maximum effort levels observed in effective designs within $5,000 of the minimum cost.

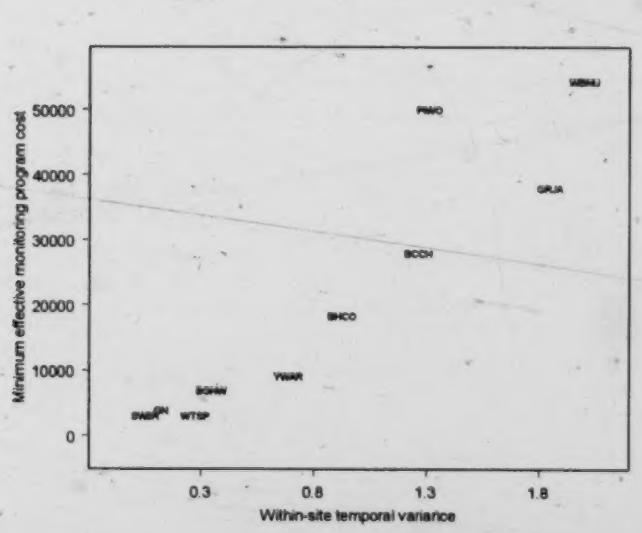

Figure 4.2. Minimum effective monitoring program cost versus within-site temporal variance (expressed as a coefficient of variation) for 20 year programs. Target codes as in Table 4.2.

respectively. Due to the insensitivity of proportions to access cost scenarios, the efficiency analysis was completed only for proportions calculated under scenario one: that of constant costs.

Table 4.3 presents the results of the efficiency analysis. The table contains power proportions divided by cost proportions for each effort parameter, excluding stations. A quotient greater than one indicates the power proportion is greater than the cost proportion, implying the effort parameter is efficient. If a quotient is less than one, the effort parameter is inefficient. Proportions were evaluated in reference to sampling designs achieving between 80 and 90% power. Investigation of other intervals, however, indicated efficiencies are consistent across different levels of power. The only inconsistencies encountered were regarding surveys for the Yellow Warbler and species richness of the ground-nesting guild. Although surveys were marginally efficient for these targets when sampling designs achieving greater than 80% were considered, they were marginally inefficient when sampling designs achieving less than 80% were evaluated.

Target	Sites	Frequency	Surveys
BCCH	2.315 (2.186, 2.457)	1.591 (1.476, 1.716)	0.742 (0.701, 0.779)
BHCO	2.138 (2.037, 2.252)	1.104 (1.014, 1.196)	0.875 (0.844, 0.904)
BGNW	1.159 (1.109, 1.211)	0.138 (0.112, 0.164)	1.146 (1.136, 1.156)
GRJA	5.121 (4.256, 6.413)	4.833 (4.043, 6.011)	-0.017 (-0.330, 0.192)
PIWO	2.117 (1.952, 2.307)	1.884 (1.722, 2.068)	0.704 (0.648, 0.753)
WBNU	4.668 (4.027, 5.545)	5.134 (4.446, 6.075)	0.141 (-0.057, 0.286)
YWAR	1.686 (1.622, 1.751)	0.356 (0.312, 0.399)	1.030 (1.013, 1.047)
GN	1.532 (1.409, 1.658)	0.374 (0.328, 0.420)	1.024 (1.006, 1.040)

Table 4.3. Relative efficiencies (and 95% confidence intervals) of effort parameters for detecting 20 year trends in various targets. An efficiency value greater than one indicates the effort parameter is efficient. Target codes as in Table 4.2.

10 Year Monitoring Programs

As observed for 20 year programs, annual program cost and the proportion of cost allocated to aerial access of sites decreased as the rate of decline in aerial access cost increased (Table 4.4). Minimum annual costs of effective programs (Table 4.5) were greater than those associated with 20 year programs. This difference was dramatic for

Scenario	No decrease	Moderate decrease	Dramatic decrease
$time_{decrease}$	0	45 seconds	90 seconds
$prop_{decrease}$	0	1.25%	2.5%
Mean annual cost	$105,207	$88,859	$77,785
Mean % of cost due to aerial access	30.6%	19.8%	9.5%

Table 4.4. Mean 10 year program annual cost and mean percentage of cost accountable to aerial access, for each aerial access cost decline scenario.

Target	Cost (scenario 1)	Cost (scenario 2)	Cost (scenario 3)	Number of sites	Sampling frequency	Number of surveys
BCCH	$358,293	$291,943	$250,906	90 (90,90)	1 (1,1)	3 (3,3)
BHCO	$162,220	$134,606	$116,538	80 (80,80)	3 (3,3)	4 (4,4)
BGNW	$24,469	$22,562	$20,847	20 (20,60)	5 (5,4)	3 (1,4)
WTSP	$8,788	$8,276	$7,765	10 (10,30)	5 (5,2)	2 (1,4)
YWAR	$61,513	$52,117	$45,405	70 (70,80)	2 (2,2)	1 (1,1)
GN	$21,362	$19,253	$17,527	40 (20,40)	3 (5,3)	1 (1,3)
SR	$5,341	$5,085	$4,829	10 (10,10)	3 (5,2)	1 (1,1)
SW	$5,341	$5,085	$4,829	10 (10,10)	5 (5,2)	1 (1,2)

Table 4.5. Least expensive effective designs for 10 year programs. Format as in Table 4.2.

temporally variable species, such as the Black-capped Chickadee, which required high sample effort. As with 20 year programs, LEED's were generally not robust (Table 4.5). Monitoring goals were not achieved by any of the simulated designs for the White-breasted Nuthatch, Pileated Woodpecker, and Gray Jay, which were therefore excluded from the efficiency analysis. All designs achieved monitoring goals for the Shannon-Weiner diversity index, which was also excluded from the efficiency analysis.

Cost partial derivatives are presented in Appendix 3. Cost partial derivative proportions were again robust to the rate of decline in aerial access cost. Average difference (and standard deviation) between proportions calculated using minimum and maximum rates of decline, across monitoring targets, were –0.0007 (0.0009), -0.0020 (0.0024), and 0.0027 (0.0033) for sampling frequency, number of surveys, and number of sites, respectively. Therefore, as with 20 year programs, efficiency analysis was restricted to the no decline in access cost scenario (Table 4.6).

Target	Sites	Frequency	Surveys
BCCH	1.607 (1.500, 1.714)	1.688 (1.560, 1.812)	0.779 (0.743, 0.814)
BHCO	1.267 (1.220, 1.333)	1.071 (0.989, 1.152)	0.956 (0.923, 0.991)
BGNW	1.500 (1.444, 1.574)	0.055 (0.005, 0.104)	1.219 (1.205, 1.232)
WTSP	1.632 (1.474, 1.797)	0.098 (0.072, 0.124)	1.440 (1.392, 1.487)
YWAR	1.189 (1.162, 1.216)	0.973 (0.886, 1.057)	1.002 (0.964, 1.041)
GN	1.460 (1.391, 1.529)	0.426 (0.355, 0.498)	1.175 (1.147, 1.204)
SR	3.000 (2.078, 3.932)	0.884 (0.573, 1.201)	0.736 (0.599, 0.874)

Table 4.6. The relative efficiencies (and 95% confidence intervals) of effort parameters for detecting 20 year trends in various targets. An efficiency value greater than one indicates the effort parameter is efficient. Target codes as in Table 4.2.

Discussion

Cost Estimates

The model and parameter estimates used to calculate cost are imprecise and the resulting cost estimates should not be regarded as accurate. In particular, the cost model accounts only for field and access costs, and not for costs associated with administration, data analysis and management, training, and report production. Therefore, cost estimates presented here underestimate true monitoring cost. The model should, however, provide a reasonable portrayal of the relative costs of the sample effort parameters. Estimation of aerial access cost was difficult due to uncertainties regarding the influences of increased efficiency as number of sites increased, and increased terrestrial access over time due to road development. Because of this uncertainty and the high proportion of total cost allocated to aerial access, program cost was calculated under three scenarios, each with a different rate of decline in aerial access. Although program cost was sensitive to the scenarios, efficiencies of sample effort parameters were robust. This supports the previous assertion that although cost estimates are likely inaccurate, estimates of the relative efficiency of sample effort parameters are reliable. Due to the insensitivity of efficiencies to the scenarios, the remainder of this discussion deals only with estimates calculated assuming no decline in aerial access costs.

Cost-effective Monitoring Program Design

Monitoring programs differed greatly in annual cost, ranging from $4,001 to $257,737 for 20 year programs and from $5,341 to $524,492 for 10 year programs. Power generally increased with program cost, although maximum cost levels did not equate to maximum power levels (e.g. Figure 4.1). Mac Nally (1997) also found that highest levels of sampling did not achieve best estimates of the maximum total density of a bird community. The imperfect relationship between cost and power implies potential for sampling design inefficiencies. The wide-range of cost for programs achieving effective data, and the limited funding available for monitoring, emphasizes the need for cost-effective monitoring program design.

Least Expensive Effective Designs

The robustness of LEED's was determined by evaluating the variability in sampling design observed within a given cost of the LEED (e.g. Millard and Lettenmaier 1986). Here, effective designs within $5,000 of the LEED cost were evaluated. This interval is small compared to the range of costs of designs achieving 90% power or greater. For example, $5,000 covers only 2.0% of the range of costs of 20 year programs achieving 90% power or greater for the Black-throated Green Warbler. In general, sample effort parameters varied greatly within the $5,000 buffer. Because of the high variability observed, effort levels associated with LEED's offer little insight and should not be used to guide monitoring program design.

Minimum costs of effective monitoring programs are useful, however, in emphasizing the sensitivity of monitoring program cost to both variance characteristics of the metric being monitored and length of the time period over which trends are to be detected. Metrics exhibiting high within-site temporal variance were more costly to monitor, reflecting the increased sample effort required (Chapter 3). Because less abundant species tend to exhibit higher within-site temporal variance (Chapter 2), effectively monitoring rare species will be costly. Detecting short-term trends will also be expensive, as indicated by the high cost of detecting trends over 10 years as compared to

20 years. Therefore, if a monitoring program is to act as an early warning system, large financial resources will be required. It is important to note that the minimum costs presented here assumed detection of a constant trend across the monitoring region. If a program is designed to monitor multiple regions, habitat types, or disturbance levels with the intent of detecting different levels of trend across classes, costs will be much greater than reported here, due to stratification.

Relative Efficiency of Effort Parameters

An alternative approach for determining cost-effective sampling strategies is to investigate the rate at which data quality increases compared to the rate at which cost increases, as effort is allocated to sample effort parameters. Such an analysis can highlight effort parameters for which increases in power are achieved relatively inexpensively. This analysis strategy is similar to marginal analysis, a commonly used tool in economics to determine optimal decisions (Baumol et al. 1985). Marginal analysis assumes effort should be allocated as long as the rate of increase in benefits (marginal benefits) is greater than or equal to the rate of increase in cost (marginal cost) associated with exertion of effort. Prior to marginal analysis, power must be converted to a monetary value to facilitate comparison of marginal costs and benefits. Benefits associated with powerful monitoring include the reduction in mitigation costs associated with early detection of ecological trends. In addition, the proactive approach to sustainable management that is facilitated by effective monitoring would enable industry to respond to government and public scrutiny (Schneider et al. 1999). This would have many strategic benefits for industry, such as increased consumer confidence, decreased government regulation, and certification. The value of these benefits must also be included in the marginal analysis. If a low-power monitoring program was implemented, and declines in biodiversity were not detected until these declines were drastic, industry would likely incur much greater cost in terms of reduced consumer confidence and increased mitigation cost than the cost of a high-power monitoring program. This suggests implementing high-power monitoring programs is an optimal economic decision

for natural resource sectors. Therefore, although not attempted here, marginal analysis could be used to highlight the value of powerful long-term ecological monitoring.

Instead of marginal analysis, this study used a cost-effectiveness analysis to identify efficient sampling strategies. This was done by comparing an effort parameter's proportional rate of increase in power to its proportional rate of increase in cost. If the proportional rate of increase in power was greater than the proportional rate of increase in cost, allocation of effort was considered efficient. To focus on the response of power and cost to sample effort within the vicinity of the 90% power goal, efficiency analysis was completed using average effort levels across designs achieving between 80 and 90% power. Efficiencies were consistent across power levels, however, indicating cost-effective sampling strategies reported here are not restricted to designs achieving between 80 and 90% power. Efficiencies were also consistent between 20 year and 10 year programs, suggesting cost-effective sampling strategies are robust to program length. As noted previously, however, shorter monitoring programs require greater overall sample effort. Because optimal effort allocation depends on the relative magnitude of variance parameters (Bernstein and Zalinski 1983), one would expect sampling efficiencies to vary across monitoring targets that exhibit different variance characteristics. The inclusion of multiple species and community metrics enabled a comparison of sampling efficiencies across different targets. This, in turn, facilitated identification of consistent sampling efficiencies across targets, and efficiencies contingent on variance characteristics of the monitoring target. Cost-effective sampling strategies suggested by the efficiency analysis are now outlined for each effort parameter, followed by an evaluation of the concordance of these sampling strategies with LEED's.

Number of Point Count Stations

Sampling the maximum number of point count stations (16) was always efficient because it required no additional cost compared to sampling fewer point count stations. Although sampling of additional stations are not free in reality, the associated cost will be minimal compared to the cost of accessing sites. Of course, power will not increase indefinitely as more stations are sampled. However, as long as additional within-site

sampling (i.e. point count stations) reduces sample error, high within-site sample effort should be a component of a large-scale monitoring program.

Number of Sites

Allocation of effort to number of sites was also efficient for all monitoring targets, suggesting that high site density is also a component of an efficient monitoring design for birds. This is consistent with Millard and Lettenmaier's (1986) finding that optimal design for benthic monitoring programs included many sites. Link et al. (1994) also reported the importance of sites, concluding from their analysis of Breeding Bird Survey data that sampling additional sites is a more efficient method of reducing total variability than repeated visits to sites within a year.

Sampling Frequency

Previous evaluation of the power of long-term avian sampling designs has assumed sampling occurs every year (e.g. Gibbs and Melvin 1997, Gibbs et al. 1998, Lewis and Gould 2000). High sampling frequency is costly, however, emphasizing the need to determine whether high allocation of effort to frequency is worthwhile. Here, simulations facilitated estimation of power for a range of sampling frequencies, in turn enabling an evaluation of the efficiency of allocating effort to sampling frequency. The importance of allocating effort to frequency was positively related with within-site temporal variance of the monitoring target. High sampling frequency is required to offset high temporal variance, resulting in allocation of effort to frequency being highly efficient for temporally variable species such as the White-breasted Nuthatch and Gray Jay. For temporally stable monitoring targets, such as the Black-throated Green Warbler, the increase in data quality achieved by increasing sampling frequency was small, making the effort parameter inefficient. Power estimates used in the efficiency analysis assumed a constant trend for the duration of the monitoring period. If, however, detection of cyclical or irregular trends is required, more frequent monitoring will be required (Scott 1998).

Number of Surveys

The efficiency of surveys was also dependent on within-site temporal variance. Surveys were inefficient for species with high temporal variance such as the Gray Jay and White-breasted Nuthatch. For these species, allocation of effort to frequency was of great importance, thereby minimizing the proportional rate of increase in power due to surveys. For some species, such as the Gray Jay, inefficiency of surveys was exacerbated by the effort parameter's low influence on within-site temporal variance (Chapter 2). In contrast, surveys were marginally efficient when monitoring targets exhibited low within-site temporal variance, such as the Black-throated Green Warbler and richness of the ground-nesting guild, reflecting the decreased importance of frequency. However, the minimum cost analysis indicated detection of trends in these metrics was achieved with low sample effort, including surveys. Therefore, repeated visits to a site within a year would not appear to be a component of an efficient long-term avian monitoring program because the power achieved by sampling multiple surveys is not great enough to offset the high expense. This supports Link et al.'s (1994) recommendation that repeated visits not be performed.

Concordance of Cost-effective Sampling Strategies with LEED's

LEED's generally utilized the sampling strategies identified as efficient. Maximum point count station density was a component of all LEED's. Number of sites was also generally high, as expected according to the efficiency analysis. The exceptions were targets for which only low sampling intensity across all effort parameters was required to achieve 90% power. LEED's generally minimized surveys, reflecting the inefficiency of the effort parameter. Although LEED's called for greater than one survey for a few species, one survey was almost always a component of designs within the $5,000 cost buffer. The exception were programs to detect 10 year trends in variable species, such as the Black-capped Chickadee, for which high sampling intensity across all effort parameters were required to achieve 90% power. Therefore, multiple surveys may be required if detection of short-term trends in variable targets is required. Sampling frequency of LEED's also agreed with strategies suggested by the efficiency analysis. Frequency was generally low for targets exhibiting low within-site temporal variance, but

higher for more variable species. Again, exceptions can be explained by the instability of LEED's.

Extending the Analysis to Additional Taxa

Results reported here suggest strategies for detecting long-term trends in bird communities. However, elements of biodiversity requiring monitoring are much more extensive (Johnson 1993, Minister of Supply and Services 1995). Monitoring multiple elements of biodiversity concurrently is undoubtedly more efficient than multiple programs, each monitoring a separate element (Lee and Hanus 1998). Recognizing the broad monitoring requirements and potential efficiencies, natural resource managers are currently developing a multi-taxa, large-scale biodiversity monitoring program for the forested region of Alberta (Schneider et al. 1999). Although such a program is theoretically more efficient than multiple specialized programs, in reality it presents a large logistical challenge. Cost-effective monitoring strategies may differ across taxa, as might minimum levels of effort required to achieve power goals. To ensure cost-effective design for multi-taxa monitoring, analyses such as the one presented here must extend beyond single taxa, facilitating a comparison, and amalgamation, of cost-effective monitoring strategies.

Summary

A cost model was developed to estimate the cost of a range of large-scale avian monitoring designs. Cost estimates were integrated with power estimates to facilitate identification of LEED's and evaluation of the efficiency of allocating effort to various sample effort parameters. LEED's were not robust, and therefore offered little insight. Efficiency analysis, on the other hand, suggested several strategies for achieving cost-effective avian monitoring design. Effort should be allocated to within-site sampling intensity, such as point count stations, due to the low expense of this activity. Repeated visits to sites within a year, however, should not be utilized due to the high associated expense. Many sites should be sampled. If the monitoring target is temporally variable,

sampling should occur frequently. Otherwise, high sample frequency is inefficient. The cost-effectiveness analysis reported here must be extended to other taxa to facilitate evaluation of monitoring strategies for large-scale biodiversity monitoring programs.

Chapter 5

General Conclusions

Introduction

This thesis has investigated cost-effective sampling design for large-scale avian monitoring. Chapter 1 provided context by emphasizing the importance of ecological monitoring and the need for cost-effective monitoring design. Chapter 2 described the structure of avian monitoring data and provided empirically based estimates of relevant parameters. In Chapter 3, these estimates were used to conduct simulations to derive estimates of the statistical power of potential sampling designs. Power estimates were integrated with cost estimates in Chapter 4, facilitating an evaluation of cost-effective sampling strategies.

In this chapter, results from previous chapters are summarized, and their implications for avian monitoring are discussed. The need to extend the analysis to additional biodiversity elements is emphasized, and the potential for doing so is explored. An alternative method for determining cost-effective monitoring design is also presented. Finally, general implications of the research are highlighted.

General Findings

In Chapter 2, eight species were selected as subjects for the study based on their variance characteristics. Generally, rare species tended to exhibit greater variability and, as exhibited by the power analysis in Chapter 3, this translated into increased sampling requirements. Due to the steep gradient in sampling requirements across species, population monitoring programs should be designed to achieve effective data for the most variable species of interest. This will ensure the full spectrum of species is adequately monitored.

Because of the increased sample effort required, effective sampling designs were expensive for rare species, compared to common species (Chapter 4). Rare species are often sensitive to environmental disturbance and are usually of conservation concern. Therefore, inclusion of rare species in population monitoring will often be desired. If the expense associated with monitoring rare species is prohibitive, careful selection of target species that are sensitive to disturbance, yet exhibit low variability, is an alternative. Species in the spatially variable category identified in Chapter 2 (i.e. Black-throated Green Warbler) may be examples of such targets, due to their habitat specificity but temporal stability.

Results suggest monitoring of community metrics may be another viable alternative. Three community metrics were included in the analysis: total species richness, species richness of the ground-nesting guild, and the Shannon-Weiner diversity index. All metrics exhibited less variance than even the most common species (Chapter 2), which translated into low sampling requirements (Chapter 3) and low program expense (Chapter 4). Two issues must be resolved before monitoring of community metrics can be unequivocally recommended, however. First, the sensitivity of community metrics to disturbance must be resolved. Community metrics are almost certainly less sensitive to disturbance than populations, and a smaller magnitude was used in community metric power analyses to account for this reduction. The magnitude of the reduction was entirely subjective, however, and may not have adequately accounted for the relative stability of community metrics. Second, it is not at all clear which community metric, or combination of metrics, should be monitored to provide meaningful information on the status of the bird community. In addition, interpretation of metrics is difficult. For example, the implications of a 10% reduction in the Shannon-Weiner diversity index are unclear. Despite these uncertainties, however, the relative low cost of monitoring community metrics is promising. At the very least, this finding warrants increased research into the use of community metrics to monitor avian communities.

Multiple sample visits to a site within a year (surveys) and multiple sample plots within a site (point count stations) were investigated as potential methods to decrease sample error. Both effort parameters substantially decreased within-site temporal variance, indicating increased within-site sample effort did reduce sample error (Chapter

2). The effect of increasing point-count stations from four to 16 and number of surveys from one to four was similar, although the relative magnitude of the effect differed across monitoring targets. As expected, both effort parameters were associated with positive increases in power (Chapter 3), although the relative performance of the parameters was difficult to interpret in the absence of cost information. This was addressed in Chapter 4, when a cost model was incorporated into the analysis. Increasing the number of surveys was far more expensive than increasing the number of point count stations, making increases in point count stations the more efficient method to decrease sample error and increase power. Therefore, increasing within-site sampling intensity is the recommended strategy to decrease sample error in large-scale avian monitoring. However, if small overall effect sizes must be detected in variable targets, such as is the case when attempting to detect trends in rare species over short time periods, multiple surveys will be required in order to achieve adequate power (Chapter 4).

In addition to sample error, data quality (power) is also negatively affected by among-unit variation, which is variation introduced by sampling only a portion of the sampling frame. Two methods of decreasing among-unit variation were investigated: increasing the number of sites, which increases the spatial coverage of the sampling frame; and increasing sampling frequency, which increases the temporal coverage of the sampling frame. The power analysis in Chapter 3 indicated both methods led to an increase in power, although, again, the magnitude of the increase varied depending on the variance characteristics of the monitoring target. Targets exhibiting high within-site temporal variance tended to respond strongly to frequency, while power to detect trends in targets exhibiting low temporal variance tended to respond strongly to number of sites. Inclusion of cost estimates in Chapter 4 revealed allocation of effort to number of sites was generally efficient across all monitoring targets, suggesting high site density is a cost-efficient avian monitoring strategy. The efficiency of sample frequency differed depending on the within-site temporal variance of the monitoring target. For temporally variable targets, the high cost of frequent monitoring was offset by dramatic increases in power. In comparison, power gains for more temporally stable targets were not large enough to justify frequent monitoring. Therefore, frequent sampling should only be used when monitoring temporally variable targets, such as rare species.

Extending the Analysis to Multiple Taxa

This thesis focused on cost-effective strategies for monitoring one component of biodiversity, the bird community. However, comprehensive monitoring of biodiversity must extend beyond birds to include diverse taxa. To take advantage of efficiencies, the ambitious goal of monitoring biodiversity should be achieved through a broad-based biodiversity monitoring program, rather than a collection of specialized programs. Realizing the advantages of a comprehensive monitoring program, natural resource managers in Alberta are currently developing the Alberta Forest Biodiversity Monitoring Program. The goal of the program is to detect changes in the biodiversity of Alberta's forests (Schneider et al. 1999). For a program of this magnitude, cost-effective monitoring becomes an issue of paramount importance.

To facilitate cost-effective design of a biodiversity monitoring program, the efficiency analysis described in this thesis should be extended to the multiple taxa of interest. A major challenge that must be overcome to achieve this goal is estimating the mean and variance parameters required to conduct power simulations. To ensure simulations are as realistic as possible, these estimates should be empirically based. Data requirements, however, may be prohibitively high (see Chapter 2, Table 6). An extensive search for data sets meeting the requirements revealed severe limitations (Rangen and Schaffer 2000). In shortest supply were data collected across multiple years; in addition to birds, sufficient data were only available for various components of terrestrial vegetation and aquatic macroinvertebrates. The lack of necessary data is not surprising, given the scarcity of long-term, spatially extensive studies. This is the void, after all, that large-scale monitoring is supposed to fill. However, if the efficiency analysis strategy used here is to be extended to multiple elements of biodiversity, a solution to present data limitations must be found.

The sensitivity analysis reported in Chapter 3 suggests some shortcuts may be possible without jeopardizing the reliability of results. Due to the moderate and low sensitivity of power to between-site variation in initial abundance and between-site variation in trend, respectively, accurate estimation of these parameters may not be essential. Instead, rough estimates could be used to facilitate power analysis, perhaps using projected minimum

and maximum variance estimates to bound power estimates. However, an empirically based solution is desired for within-site temporal variance, due to the sensitivity of power to this parameter. A potential strategy is to use estimates of the temporal variability for a range of local plant and animal populations presented in Gibbs et al. (1998). To achieve these estimates, Gibbs et al. (1998) obtained all count series of animal and plant populations at least five years in length from major ecology journals from 1940 onwards. Variance of the residuals of a linear regression of counts against time was used as the estimate. No selection criteria were used when selecting data sets beyond the minimum length criterion, and therefore the resulting estimates are likely very inaccurate. Nevertheless, these rough estimates could be used to compare sampling requirements across biodiversity elements. However, the estimates do not facilitate an evaluation of the influence of within-site sample effort on power. Results reported here indicate that within-site sample effort can have a dramatic influence on power, thus this component of effort should not be ignored. Due to the lack of available data, alternative methods will be required to identify cost-effective levels of within-site sample effort.

The solutions presented above are less than ideal, but may be pragmatic. At the very least, they should facilitate a rough comparison of cost-effective strategies across elements of biodiversity, thereby guiding the development of strategies that consider the diverse sampling requirements of potential biodiversity monitoring targets.

Optimal Experimental Design

In this study, the problem of designing a cost-effective monitoring program was interpreted as achieving high power to detect trend at low cost. Although power to detect trend is an appropriate index of data quality if the monitoring goal is detection of change (Chapter 3), power analysis is by no means the only suitable tool for evaluating efficient design. An alternative framework for identifying cost-effective sampling design is to view it as a constrained optimization problem (Walter and Pronzato 1989). Design constraints, such as cost, range of treatments, and length of the experiment, define the set of possible designs. Given these constraints, the goal is to select the design that achieves maximum information. This approach to identifying effective sampling designs is known

as optimal experimental design (e.g. Silvey 1980, Pazman 1986). Given a model for the phenomenon of interest, the optimal design is identified through evaluation of the inverse of the error covariance matrix of the best linear unbiased estimator. This matrix, known as the Fisher information matrix, is appropriate for evaluation because it contains information about measurement errors and parameter sensitivities, allowing the quality of parameter estimates achieved by a design to be quantified (Versyck et al. 1999). The optimal design is that which maximizes an optimality criterion of the Fisher information matrix. For example, when parameter estimation is the goal, as is the case with trend detection, the design that maximizes the determinant of the Fisher information matrix is often selected as the optimal design. This optimality criterion, known as D-optimality, is intuitively pleasing, as it equates to minimizing the confidence regions around parameter estimates (Walter and Pronzato 1989). As is the case with power analysis, the optimal design framework explained here requires specification of a model for the phenomenon of interest. However, optimal experimental design theory can also be used to select optimal designs when the goal is discriminating between candidate models, although an alternative optimality criterion is required (e.g. Atkinson and Fedorov 1975).

During this cursory treatment of optimal experimental design, the term monitoring was replaced with experiment. In part, this is in keeping with terminology in the optimal design literature. It also emphasizes a critical difference between the monitoring philosophies implicit in the two strategies for identifying efficient designs. The evaluation of cost-effective monitoring design reported in this thesis has assumed the monitoring goal is regional detection of long-term trend using fixed sites randomly located within the region of interest. There is no consideration of treatments at the sites, such as disturbance or habitat type. Such a program is not designed to answer any particular question, such as the effect of a disturbance type on biodiversity, but instead is designed to detect broad-scale changes in biodiversity, regardless of the cause. Although this type of monitoring plays an important role, essentially acting as an alarm bell, it does not maximize the potential information gained using monitoring funds. Optimal experimental design takes a very different approach. Treatment levels at potential sites are considered when selecting the design. The number of sites, and their locations, is selected to achieve the best possible estimate of the parameter of interest. Therefore, if

the response of biodiversity to seismic activity is of interest, sites will be selected in regions of high and low seismic activity to maximize the contrast. In doing so, the monitoring design takes advantage of current management practices being applied to the landscape to evaluate the effect of disturbances on biodiversity. This mimics the strategy of treating current management practices as experiments that is a vital component of adaptive management (Walters 1986). Optimal experimental design therefore has greater potential for providing estimates of the effect of specific management policies on biodiversity, thereby facilitating adaptive management.

As outlined, the two design strategies produce different styles of monitoring programs, and each style has its weaknesses. The cost-effectiveness analysis used in this thesis identifies monitoring programs capable of detecting regional trends. This type of program does not maximize information gain, however, and is unlikely to be able to link biodiversity trends with causal factors, thereby restricting the translation of monitoring findings into policy. On the other hand, the long-term utility of monitoring programs designed to evaluate the effect of specific disturbances, such as programs identified using optimal experimental design, may be limited. Data from a program designed to answer today's specific questions is unlikely to address tomorrow's environmental dilemmas. Instead of relying solely on either strategy, a combination should be used to take advantage of the strengths of both approaches. This would consist of designs that optimally track the effects of specific disturbances, superimposed on a large-scale design that achieves regional trend estimates with high power.

General Implications

Effective monitoring programs must provide reliable estimates of change. As illustrated by the power analyses presented in this thesis, reliable trend estimates are not a guaranteed product, but instead are dependent on the sampling design used and the monitoring target. *A priori* evaluation of the feasibility of meeting monitoring targets is essential. Equally important, however, is program cost. Programs spanning a wide-range of costs achieved reliable trend estimates, indicating great potential for sampling design efficiencies. Indeed, comparison of cost and power gains associated with sample effort

parameters revealed general strategies for achieving cost-effective data. To ensure limited resources are not wasted, *a priori* evaluation of efficient monitoring strategies is also essential.

Monitoring of biodiversity is required if responsible management of the environment is to be achieved. Monitoring is a costly endeavor, however, and limited funds are available. To ensure adequate monitoring is achieved under fiscal constraints, analyses that consider both the quality of monitoring data and the efficiency of sampling strategies are a vital component of monitoring program design.

Literature Cited

Alberta Forest Conservation Strategy Steering Committee. 1997. Alberta Forest Conservation Strategy: A New Perspective on Sustaining Alberta's Forests. Alberta Land and Forest Service, Alberta Environmental Protection, Edmonton, Alberta.

Alberta Environmental Protection. 1994. Natural Regions and Subregions of Alberta: Summary. Alberta Environmental Protection, Edmonton, Alberta.

Alberta Environmental Protection. 1998. The Alberta Forest Legacy: Implementation Framework for Sustainable Forest Management. Pub. No.: **I/689**. Alberta Environmental Protection, Edmonton, Alberta.

Atkinson, A. C. and V. V. Fedorov. 1975. Optimal design: experiments for discriminating between several models. Biometrika **62**: 289-303.

Baillie, S. R. 1991. Monitoring terrestrial breeding bird populations. Pages 112-132 *in* B. Goldsmith, editor. Monitoring for Conservation and Ecology. Chapman and Hall, New York.

Barker, R. J., J. R. Sauer, and W. A. Link. 1993. Optimal allocation of point-count sampling effort. The Auk. **110(4)**: 752-758

Baumol, W. J., A. S. Blinder, and W. M. Scarth. 1985. Economics, Principles and Policy. Academic Press, Canada.

Bernstein, B. B. and J. Zalinski. 1983. An optimum sampling design and power tests for environmental biologists. Journal of Environmental Management **16**: 35-43.

Bibby, C. J., N. D. Burgess, and D. A. Hill. 1993. Bird Census Techniques. Academic Press, San Diego, California.

Bull, E. L. and J. A. Jackson. 1995. Pileated Woodpecker. The Birds of North America **48**.

Burbidge, A. A. 1991. Cost constraints on surveys for nature conservation. Pages 3-6 *in* C. R. Margules and M. P. Austin, editors. Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Australia.

Caughley, G. and A. R. E. Sinclair. 1994. Wildlife Ecology and Management. Blackwell Science, Cambridge, Massachusetts.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, USA.

Committee on Environment and Natural Resources. 1996. Integrating Environmental Monitoring and Research in the mid-Atlantic Region. Committee on Environment and Natural Resources, National Science and Technology Council, Washington D.C. (Available on the internet at http://www.epa.gov/cludygwb/Pubs/marrpt.pdf)

Committee on Environment and Natural Resources. 1997. Integrating the Nation's Environmental Monitoring and Research Networks and Programs: a Proposed Framework. Environmental Monitoring Team, Committee on Environment and Natural Resources, National Science and Technology Council, Washington D.C. (Available on the internet at http://www.epa.gov/cludygxb/Pubs/framewrk.pdf)

Croonquist, M. J. and R. P. Brooks. 1991. Use of avian and mammalian guilds as indicators of cumulative impacts in riparian-wetland areas. Environmental Management **15**: 701-714.

Dawson, D. G. 1981. Counting birds for a relative measure (index) of density. Studies in Avian Biology **6**: 12-16.

Dixon, M, A. R. Olsen, and B. M. Kahn. 1998. Measuring trends in ecological resources. Ecological Applications **8(2)**: 225-227.

Droege, S. 1990. The North American Breeding Bird Survey. Pages 1-4 *in* J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report **90(1)**. U.S. Fish and Wildlife Service, Washington D.C.

Dunn, E. H., M. D. Cadman, and J. B. Falls. 1993. Monitoring Bird Populations: the Canadian Experience. Occasional Paper **95**. Canadian Wildlife Service, Canada.

Fairweather, P. G. 1991. Statistical power and design requirements for environmental monitoring. Australian Journal of Marine Freshwater Research **42**: 555-67.

Farr, D. R., S.E. Franklin, E.E. Dixon, G. Scrimgeour, S. Kendall, P. Lee, S. Hanus, N.N. Winchester and C.C. Shank. 1999. Monitoring Forest Biodiversity in Alberta: Program Framework. Alberta Forest Biodiversity Monitoring Program Technical Report **3**. Draft Report.

Furness, R. W., J. J. D. Greenwood, and P. J. Jarvis. 1991. Can birds be used to monitor the environment? Pages 1-41 *in* R. W. Furness and J. J. D. Greenwood, editors. Birds as Monitors of Environmental Change. Chapman and Hall, London, England.

Gibbs, J. P., S. Droege and P. Eagle. 1998. Monitoring populations of plants and animals. BioScience **48**: 935-940.

Gibbs, J. P. and S. M. Melvin. 1997. Power to detect trends in waterbird abundance with call-response surveys. Journal of Wildlife Management **61(4)**: 1262-1267.

Geissler, P. H., and W. A. Link. 1988. Bias of animal population trend estimates. Pages 755-759 *in* E. J. Wegman, D. T. Gantz and J. J. Miller, editors. Computing Science and Statistics; Proceedings of the 20th Symposium on the Interface; Fairfax, Virginia; April 1988. American Statistical Society, Alexandria, Louisiana.

Geissler, P. H. and J. R. Sauer. 1990. Topics in route-regression analysis. Pages 54-57 *in* J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report **90(1)**. U.S. Fish and Wildlife Service, Washington, D.C.

Hagan, J. M. 1992. Conservation biology when there is no crisis – yet. Conservation Biology **6**: 475-476.

Hassell, M. P. 1989. Workshop on environmental priorities. Biologist **36**: 275-280.

Hayes, J. P. and R. J. Steidl. 1997. Statistical power analysis and amphibian population trends. Conservation Biology **11(1)**: 273-275.

Heywood, V. H. and I. Baste. 1995. Introduction. Pages 3-19 *in* V. H. Heywood and R. T. Watson, editors. Global Biodiversity Assessment. Cambridge University Press, Cambridge, Great Britain.

Hilborn, R. and M. Mangel. 1997. The Ecological Detective. Princeton University Press, Princeton, New Jersey.

Hinds, W. T. 1984. Towards monitoring of long-term trends in terrestrial ecosystems. Environmental Conservation **11**: 11-17.

Johnson, S. P. 1993. The Earth Summit: The United Nations Conference on Environment and Development. Graham and Trotman/Martinus Nijhoff, London.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries **6**: 21-27.

Karr, J. R. and E. W. Chu. 1997. Biological Monitoring and Assessment: Using Multimetric Indexes Effectively. **EPA 235-R97-001**. University of Washington, Seattle.

Krinsky, I. and A. L. Robb. 1986. On approximating the statistical properties of elasticities. The Review of Economics and Statistics **68**: 715-719.

Lee, P. and S. Hanus. 1998. Criteria and Indicators for Monitoring Biodiversity in Alberta's Forests: Part I: Review of Legislation, Policies, External Agreements, and Programs. Alberta Environmental Protection. (Available on the Internet at http://www.gov.ab.ca/~env/resrched.html.)

Lesica, P. and B. M. Steele. 1996. A method for monitoring long-term population trends: an example using rare arctic-alpine plants. Ecological Applications **6**: 879-887.

Lewis, S. A. and W. R. Gould. 2000. Survey effort effects on power to detect trends in raptor migration counts. Wildlife Society Bulletin **28**: 317-329.

Link, W. A., R. J. Barker, J. R. Sauer, and S. Droege. 1994. Within-site variability in surveys of wildlife populations. Ecology **75(4)**: 1097-1108.

Link, W. A. and J. S. Hatfield. 1990. Power calculations and model selection for trend analysis: a comment. Ecology **71**: 1217-1220.

Lowther, P. E. 1993. Brown-headed Cowbird. The Birds of North America, **47**.

Mac Nally, R. 1997. Monitoring forest bird communities for impact assessment: the influence of sampling intensity and spatial scale. Biological Conservation **82**: 355-367.

Magurran, A. E. 1988. Ecological Diversity and Its Measurement. Princeton University Press, Princeton, New Jersey.

Margalef, R. 1958. Information theory in ecology. General Systems **3**: 36-71.

MathSoft. 1999. S-Plus 2000 Guide to Statistics, Volume 1. Data Analysis Products Division, MathSoft, Seattle, WA.

Millard, S. P. and D. P. Lettenmaier. 1986. Optimal design of biological sampling programs using the analysis of variance. Estuarine, Coastal and Shelf Science **22**: 637-656.

Minister of Supply and Services Canada. 1995. Canadian Biodiversity Strategy: Canada's Response to the Convention on Biological Diversity. Cat. No. En21-134/1995E.

Moors, A. K. 1993. Towards an Avian Index of Biotic Integrity for Lakes. Masters thesis, University of Maine.

Myers, R. H. 1990. Classical and Modern Regression with Applications. Duxbury Press, Belmont, California.

National Research Council. 1995. Review of EPA's Environmental Monitoring and Assessment Program: Overall Evaluation. National Academy Press, Washington, DC.

Natural Resources Canada. 1998. National Forest Strategy 1998-2003. Sustainable Forests: a Canadian Commitment. Natural Resources Canada, Canadian Forest Service, Ottawa, Canada.

Pazman, A. 1986. Foundations of Optimum Experimental Design. D. Reidel Publishing Company, Czechoslovakia.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences **47**: 2-15.

Pimm, S. L. 1991. The Balance of Nature? Ecological Issues in the Conservation of Species and Communities, University of Chicago Press, Chicago.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, New York.

Rangen, S. and W. Schaffer. 2000. Project 1 Progress Report. Unpublished report produced for the Alberta Forest Biodiversity Monitoring Program.

Root, R. B. 1967. The niche exploitation pattern of the blue-gray gnatcatcher. Ecological Monographs **103**: 1-34.

Samways, M. J., N. E. Stork, J. Cracraft, H. A. C. Eeley, M. Foster, G. Lund, and C. Hilton-Taylor. 1995. Scales, planning and approaches to inventorying and monitoring. Pages 475-517 *in* V. H. Heywood and R. T. Watson, editors. Global Biodiversity Assessment. Cambridge University Press, Cambridge, Great Britain.

Schluter, D. and R. E. Ricklefs. 1993. Species diversity: an introduction to the problem. Pages 1-10 *in* R. E. Ricklefs and D. Schluter, editors. Species Diversity in Ecological Communities. The University of Chicago Press, Chicago.

Schmiegelow, F. K., C. S. Machtans, and S. J. Hannon. 1997. Are boreal birds resilient to forest fragmentation? An experimental study of short-term community responses. Ecology **78**: 1914-1932.

Schneider, R., P. Lee, C. Shank, D. Farr, B. Stelfox. 1999. Conceptual framework and rationale for monitoring forest biodiversity in Alberta. Pages 1-45 *in* D.R. Farr, S.E. Franklin, E.E. Dixon, G. Scrimgeour, S. Kendall, P. Lee, S. Hanus, N.N. Winchester and C.C. Shank. Monitoring Forest Biodiversity in Alberta: Program Framework. Alberta Forest Biodiversity Monitoring Program Technical Report **3**. Draft Report.

Scott, C. T. 1998. Sampling methods for estimating change in forest resources. Ecological Applications **8(2)**: 228-233.

Silvey, S. D. 1980. Optimal Design: an Introduction to the Theory for Parameter Estimation. Chapman and Hall, London, England.

Sobol, I. M. 1975. The Monte Carlo Method. Mir Publishers, Moscow.

Spellerberg, I. F. 1991. Monitoring Ecological Change. Cambridge University Press, Cambridge, England.

Steele, B. B., R. L. Bayn, Jr. and C. Val Grant. 1984. Environmental monitoring using populations of birds and small mammals: analyses of sampling effort. Biological Conservation **30**: 157-172.

Steidl, R. J., J. P. Hayes, and E. Schauber. 1997. Statistical power analysis in wildlife research. Journal of Wildlife Management **61(2)**: 270-279.

Stork, N. E. and M. J. Samways. 1995. Inventorying and monitoring: executive summary. Pages 457-458 *in* V. H. Heywood and R. T. Watson, editors. Global Biodiversity Assessment. Cambridge University Press, Cambridge, Great Britain.

Stork, N. E., M. J. Samways, and D. A. Bryant. 1995. Why inventory and monitor biodiversity? Pages 461-475 *in* V. H. Heywood and R. T. Watson, editors. Global Biodiversity Assessment. Cambridge University Press, Cambridge, Great Britain.

Taylor, B. L. and T. Gerrodette. 1993. The uses of statistical power in conservation biology: the Vaquita and Northern Spotted Owl. Conservation Biology **7**: 489-500.

Thomas, R. 1998. The Boreal Forest Natural Region of Alberta. Alberta Environmental Protection, Edmonton, Alberta.

Thomas, L. 1997a. Evaluation of Statistical Methods for Estimating Long-term Population Change from Extensive Wildlife Surveys. Ph. D. thesis, University of British Columbia, Vancouver, British Columbia.

Thomas, L. 1997b. USNBS Route Regression Program.

Thompson, W. L., G. C. White, C. Gowan. 1998. Monitoring Vertebrate Populations. Academic Press, San Diego, California.

Tyler Miller, G. Jr. 1993. Environmental Science: Sustaining the Earth. Wadsworth Publishing Company, Belmont, California.

United States Environmental Protection Agency. 2000. Environmental Monitoring and Assessment Program Homepage. http://www.epa.gov/emap/.

Urquhart, N. S., S. G. Paulsen and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications **8**: 246-257.

Usher, M. B. 1989. Scientific aspects of nature conservation in the United Kingdom. Journal of Applied Ecology **26**: 813-824.

USGS Patuxent Wildlife Research Center. 2000. BBS Physiographic Strata Map, (http://www.mp2-pwrc.usgs.gov/bbs/bbs/strata_map_small.htm).

USGS Patuxent Wildlife Research Center. 2000. North American Breeding Bird Survey Internet data set, (http://www.mp2-pwrc.usgs.gov/bbs/retrieval/).

Verner, J. 1988. Optimizing the duration of point counts for monitoring trends in bird populations. Res. Note **PSW-395**. Pacific Southwest Forest and Range Experiment Station, Forest Service, U. S. Department of Agriculture, Berkeley, California.

Versyck, K. J., K. Bernaerts, A. H. Geeraerd, and J. F. van Impe. 1999. Introducing optimal experimental design in predictive modeling: a motivating example. International Journal of Food Microbiology **51**: 39-51.

Walter, E. and L. Pronzato. 1990. Qualitative and quantitative experiment design for phenomenological models – a survey. Automatica **26**: 195-213.

Walters, C. 1986. Adaptive Management of Renewable Resources. Macmillan Publishing Company, New York.

Walters, C. J., and C. S. Holling. 1990. Large-scale management experiments and learning by doing. Ecology **71**: 2060-2068.

Zielinski, W. J. and H. B. Stauffer. 1996. Monitoring *Martes* populations in California: survey design and power analysis. Ecological Applications **6(4)**: 1254-1267.

Appendix 1

Parameter Estimates

Target	τ_1^2 (CV)											
	4 stations				9 stations				16 stations			
	1 survey	2 surveys	3 surveys	4 surveys	1 survey	2 surveys	3 surveys	4 surveys	1 survey	2 surveys	3 surveys	4 surveys
BCCH	2.389 2.327	2.274 2.188	2.020 2.003	1.903 1.903	2.185 2.120	1.922 1.834	1.732 1.657	1.570 1.570	2.035 1.887	1.740 1.610	1.455 1.385	1.264 1.264
BGNW	1.219 1.073	0.806 0.809	0.734 0.680	0.615 0.615	0.773 0.718	0.544 0.514	0.470 0.442	0.407 0.407	0.605 0.562	0.442 0.427	0.367 0.381	0.348 0.348
BHCO	1.978 2.096	1.747 1.887	1.689 1.730	1.629 1.629	1.659 1.839	1.464 1.516	1.448 1.349	1.229 1.229	1.422 1.578	1.399 1.314	1.331 1.102	0.925 0.925
GRJA	1.849 2.391	1.849 2.320	1.849 2.284	2.326 2.326	1.849 2.387	1.849 2.269	1.849 2.257	2.256 2.256	1.549 2.296	1.549 1.937	1.549 1.830	1.848 1.848
PIWO	2.449 2.406	2.449 2.322	2.449 2.260	2.064 2.064	2.449 2.410	2.449 2.269	2.449 2.131	1.857 1.857	2.449 2.256	2.449 1.977	2.449 1.607	1.314 1.314
WBNU	2.449 2.449	2.449 2.449	2.499 2.449	1.999 1.999	2.449 2.387	2.449 2.314	1.999 2.299	2.224 2.224	2.449 2.366	2.449 2.202	1.549 2.127	1.999 1.999
WTSP	0.671 0.615	0.468 0.460	0.406 0.391	0.357 0.357	0.573 0.513	0.383 0.383	0.353 0.340	0.313 0.313	0.484 0.439	0.350 0.346	0.327 0.304	0.276 0.276
YWAR	1.801 1.525	1.358 1.250	1.158 1.096	0.991 0.991	1.462 1.277	1.087 1.041	1.036 0.890	0.795 0.795	1.218 1.138	0.964 0.894	0.990 0.768	0.688 0.688
SR	0.206 0.234	0.153 0.162	0.143 0.139	0.122 0.122	0.141 0.200	0.100 0.139	0.099 0.111	0.105 0.105	0.117 0.186	0.089 0.123	0.092 0.099	0.086 0.086
GN	0.329 0.321	0.231 0.249	0.230 0.216	0.185 0.185	0.247 0.279	0.168 0.209	0.183 0.172	0.160 0.160	0.247 0.232	0.158 0.201	0.164 0.162	0.124 0.124
SW	0.100 0.129	0.069 0.077	0.057 0.057	0.046 0.046	0.062 0.080	0.044 0.054	0.035 0.043	0.033 0.033	0.047 0.069	0.029 0.046	0.023 0.033	0.026 0.026

Table A1.1. Within-site temporal variance estimates (τ_1^2), expressed as coefficients of variation. Two estimates are provided for each species at each within-site effort level. The first estimate was generated using fixed timing of surveys, while the second estimate was generated using variable timing of surveys. The target codes are: BCCH = Black-capped Chickadee, BGNW = Black-throated Green Warbler, BHCO = Brown-headed Cowbird, GRJA = Gray Jay, PIWO = Pileated Woodpecker, WBNU = White-breasted Nuthatch, WTSP = White-throated Sparrow, and YWAR = Yellow Warbler, SR = species richness, GN = richness of the ground-nesting guild, and SW = Shannon-Weiner diversity index.

Target	BCCH	BGNW	BHCO	GRJA	PIWO	WBNU	WTSP	YWAR	SR	GN	SW
τ_2^2 (CV)	0.793	2.254	1.257	0.666	1.410	0.491	0.557	0.716	0.222	0.283	0.104
τ_3^2 (CV)	0.487	0.167	0.404	0.301	0.290	0.449	0.139	0.188	0.038	0.050	0.064

Table A1.2. Between-site initial variance (τ_2^2) and between-site trend variance (τ_3^2) estimates, expressed as coefficients of variation. Target codes as in Table A1.1.

Target	α_0											
	4 stations				9 stations				16 stations			
	r=1	r=2	r=3	r=4	r=1	r=2	r=3	r=4	r=1	r=2	r=3	r=4
BCCH	0.299	0.163	0.163	0.153	0.380	0.259	0.258	0.281	0.500	0.347	0.417	0.458
	0.291	0.214	0.163	0.153	0.377	0.341	0.287	0.281	0.572	0.525	0.457	0.458
BGNW	1.694	1.729	1.483	1.579	3.880	3.963	3.392	3.639	7.139	7.181	6.185	6.563
	1.598	1.574	1.575	1.579	3.638	3.586	3.652	3.639	6.420	6.561	6.566	6.563
BHCO	0.293	0.228	0.181	0.172	0.463	0.407	0.333	0.365	0.806	0.667	0.556	0.604
	0.294	0.205	0.183	0.172	0.461	0.394	0.374	0.365	0.718	0.637	0.601	0.604
GRJA	0.389	0.194	0.130	0.108	0.389	0.194	0.130	0.125	0.500	0.250	0.167	0.229
	0.334	0.186	0.140	0.108	0.349	0.215	0.154	0.125	0.440	0.364	0.286	0.229
PIWO	0.250	0.125	0.083	0.077	0.375	0.188	0.125	0.117	0.417	0.208	0.139	0.188
	0.208	0.118	0.084	0.077	0.252	0.158	0.125	0.117	0.336	0.206	0.194	0.188
WBNU	0.167	0.083	0.056	0.049	0.167	0.083	0.083	0.066	0.167	0.083	0.111	0.083
	0.182	0.092	0.063	0.049	0.211	0.115	0.080	0.066	0.236	0.134	0.099	0.083
WTSP	4.426	4.632	4.972	5.076	9.926	10.44	11.16	11.38	17.33	18.22	19.42	19.91
	5.108	5.041	5.092	5.076	11.27	11.35	11.37	11.38	20.09	19.89	19.90	19.91
YWAR	0.637	0.757	0.798	0.922	1.273	1.581	1.671	1.998	1.944	2.375	2.602	3.146
	0.997	0.928	0.923	0.922	1.992	1.979	1.988	1.998	3.085	3.133	3.144	3.146
SR	11.81	15.29	17.24	20.43	16.73	20.57	23.14	26.77	20.39	24.58	27.44	31.64
	11.19	15.63	18.43	20.43	15.84	21.05	24.44	26.77	19.34	25.63	29.16	31.64
GN	3.898	4.588	4.903	5.731	4.972	5.556	5.954	7.139	5.611	6.222	6.722	8.306
	3.635	4.668	5.282	5.731	4.700	5.881	6.580	7.139	5.437	6.848	7.628	8.306
SW	3.179	3.417	3.486	3.640	3.558	3.689	3.742	3.855	3.733	3.814	3.860	3.956
	3.062	3.409	3.562	3.640	3.459	3.705	3.798	3.855	3.641	3.838	3.918	3.956

Table A1.3. Mean abundance estimates for species and community metrics selected as targets for subsequent sample design evaluation. Two estimates are provided for each species at each within-site effort level. The first estimate was generated using fixed timing of surveys, while the second estimate was generated using variable timing of surveys. r indicates number of surveys. Target codes as in Table A1.1.

Appendix 2

Power Model Partial Derivatives

Sample effort parameter	Partial derivative (rate of increase in power)
Number of sites (s)	$(a_1+a_5f+a_8p+a_7r)*(-exp(-1*(a_0+a_1s+a_2f+a_3p+a_4r+a_5sf+a_6sp+a_7sr+a_8fr+a_9fp+a_{10}pr)))$
Number of sample years (f)	$(a_2+a_5s+a_9r+a_9p)*(-exp(-1*(a_0+a_1s+a_2f+a_3p+a_4r+a_5sf+a_6sp+a_7sr+a_8fr+a_9fp+a_{10}pr)))$
Number of point count stations (p)	$(a_3+a_6s+a_9f+a_{10}r)*(-exp(-1*(a_0+a_1s+a_2f+a_3p+a_4r+a_5sf+a_6sp+a_7sr+a_8fr+a_9fp+a_{10}pr)))$
Number of surveys (r)	$(a_4+a_7s+a_8f+a_{10}p)*(-exp(-1*(a_0+a_1s+a_2f+a_3p+a_4r+a_5sf+a_6sp+a_7sr+a_8fr+a_9fp+a_{10}pr)))$

Table A2.1. Partial derivatives of the power model for 20 year programs (Equation 3.1).

Sample effort parameter	Partial derivative (rate of increase in power)
Number of sites (s)	$(a_1+a_4f+a_5r)*(-exp(-1*(a_0+a_1s+a_2f+a_3r+a_4sf+a_5sr+a_6fr)))$
Number of sample years (f)	$(a_2+a_4s+a_6r)*(-exp(-1*(a_0+a_1s+a_2f+a_3r+a_4sf+a_5sr+a_6fr)))$
Number of surveys (r)	$(a_3+a_5s+a_6f)*(-exp(-1*(a_0+a_1s+a_2f+a_3r+a_4sf+a_5sr+a_6fr)))$

Table A2.2. Partial derivatives of the power model for 10 year programs (Equation 3.2).

Appendix 3

Cost Model Partial Derivatives

Sample effort parameter	Partial derivative (rate of increase in cost)
Number of sites (s)	$f*r*(crew_{cost}/crew_{sites}+heli_{prop}*heli_{time}*heli_{cost})+initial_{cost}+heli_{prop}*heli_{time}*heli_{cost}$
Number of sample years(f)	$s*r*(crew_{cost}/crew_{sites}+heli_{prop}*heli_{time}*heli_{cost})$
Number of surveys (r)	$s*f*(crew_{cost}/crew_{sites}+heli_{prop}*heli_{time}*heli_{cost})$

Table A3.1. Partial derivatives of the cost model. Terms in the partial derivatives are defined in Equations 4.1 and 4.2.

Target	Number of sites			Sampling frequency			Number of rounds		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
BCCH	$33,705	$26,996	$23,767	$112,166	$92,961	$78,886	$428,217	$354,896	$301,163
BHCO	$31,695	$25,609	$22,415	$99,269	$82,578	$69,815	$370,145	$307,912	$260,322
BGNW	$28,436	$24,082	$20,474	$32,770	$27,829	$23,047	$162,170	$137,716	$114,054
GRJA	$44,092	$34,861	$30,984	$125,960	$103,844	$88,587	$653,086	$538,420	$459,313
PIWO	$45,036	$35,694	$31,666	$137,528	$113,509	$96,723	$580,967	$479,501	$408,592
WBNU	$55,798	$44,233	$39,242	$120,069	$99,140	$84,444	$816,472	$674,150	$574,222
YWAR	$30,881	$26,118	$22,190	$39,886	$33,865	$28,052	$149,988	$127,345	$105,486
GN	$10,930	$9,489	$8,211	$12,639	$10,764	$8,889	$88,472	$75,347	$62,222

Table A3.2. Rates of increase in total cost with sample effort parameters for 20 year programs. Target codes as in Table 3. The target codes are: BCCH = Black-capped Chickadee, BGNW = Black-throated Green Warbler, BHCO = Brown-headed Cowbird, GRJA = Gray Jay, PIWO = Pileated Woodpecker, White-breasted Nuthatch, YWAR = Yellow Warbler, GN = species richness of the ground-nesting guild.

Target	Number of sites			Sampling frequency			Number of rounds		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
BCCH	$36,741	$26,303	$20,645	$282,153	$234,000	$203,997	$978,132	$811,201	$707,189
BHCO	$29,169	$21,314	$16,925	$276,842	$231,665	$202,916	$677,141	$566,641	$496,322
BGNW	$15,857	$13,779	$12,267	$58,810	$53,045	$48,216	$218,085	$196,709	$178,799
WTSP	$9,666	$8,951	$8,307	$25,278	$23,403	$21,528	$37,917	$35,104	$32,292
YWAR	$19,507	$15,612	$13,170	$156,816	$136,460	$121,920	$351,080	$305,507	$272,955
GN	$18,012	$16,074	$14,552	$54,838	$50,077	$45,778	$133,805	$122,188	$111,699
SR	$6,507	$6,070	$5,675	$12,639	$11,701	$10,764	$44,236	$40,955	$37,674

Table A3.3. Rates of increase in total cost with sample effort parameters for 10 year programs. Target codes as in Table 3. Target codes as in Table A3.2, and WTSP = White-throated Sparrow, SR = species richness.